Research Needs for Biomedical Imaging

Michael W. Vannier
University of Iowa

- Includes work done at:

Biomedical Imaging Program
National Cancer Institute

Outline

- Needs assessment
- Variation in biological systems
- Specificity (of imaging)
- Questions that involve imaging = motivation
- Needs; Barriers; Metrics
- Potential Solutions
- Conclusion

Needs assessment

- Published literature
- Conferences
- Planning meetings
- National priorities, including other agencies
- Industry and investment community
- Research opportunities (unsolicited proposals)

Variation

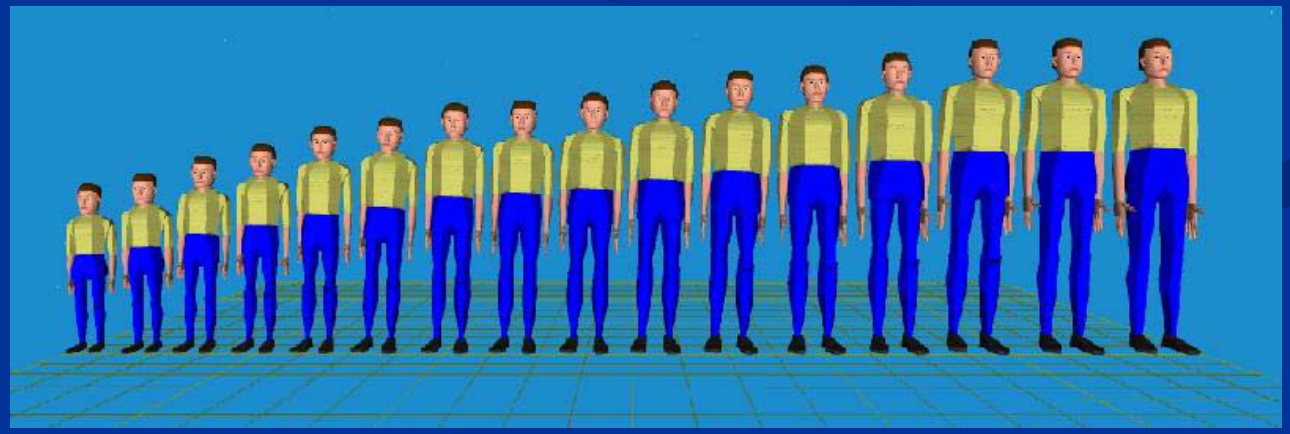
- Normal
  - Within individual differences (growth & development; response to environment or intervention)
  - Between individual differences (taxonomy: class, group, species)
    - Within species (and sub-species = defined population)
    - Between species
- Abnormal
  - Disease-related
  - Treatment-related
  - Environmental
- Measurement system
  - Precision and accuracy
- Lack of "gold standard" criterion reference object
- Temporal
  - Short term
  - Long term

Human Variation; Growth & Development

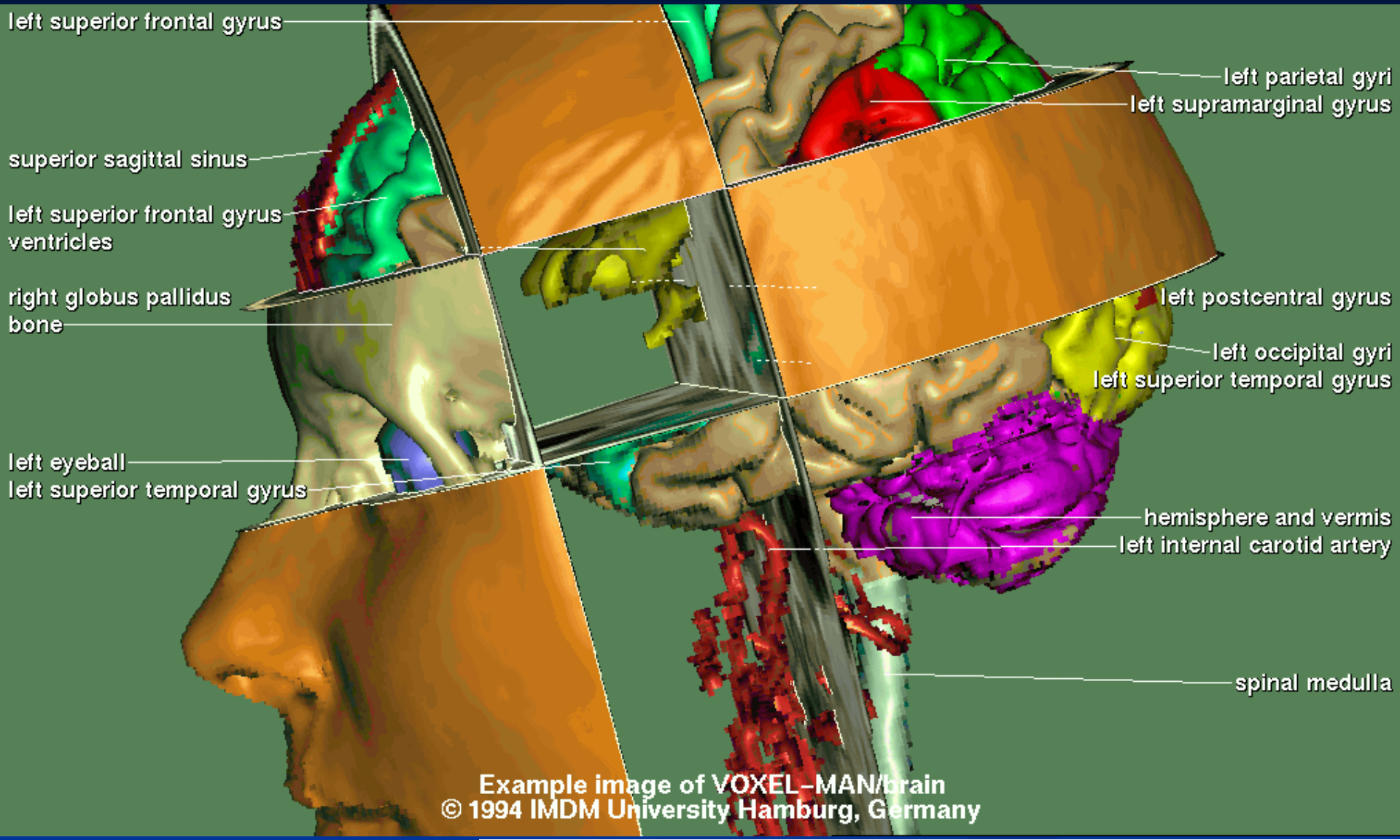
Population:

normal variants,
sexual dimorphism,
racial/ethnic character,
heritable differences

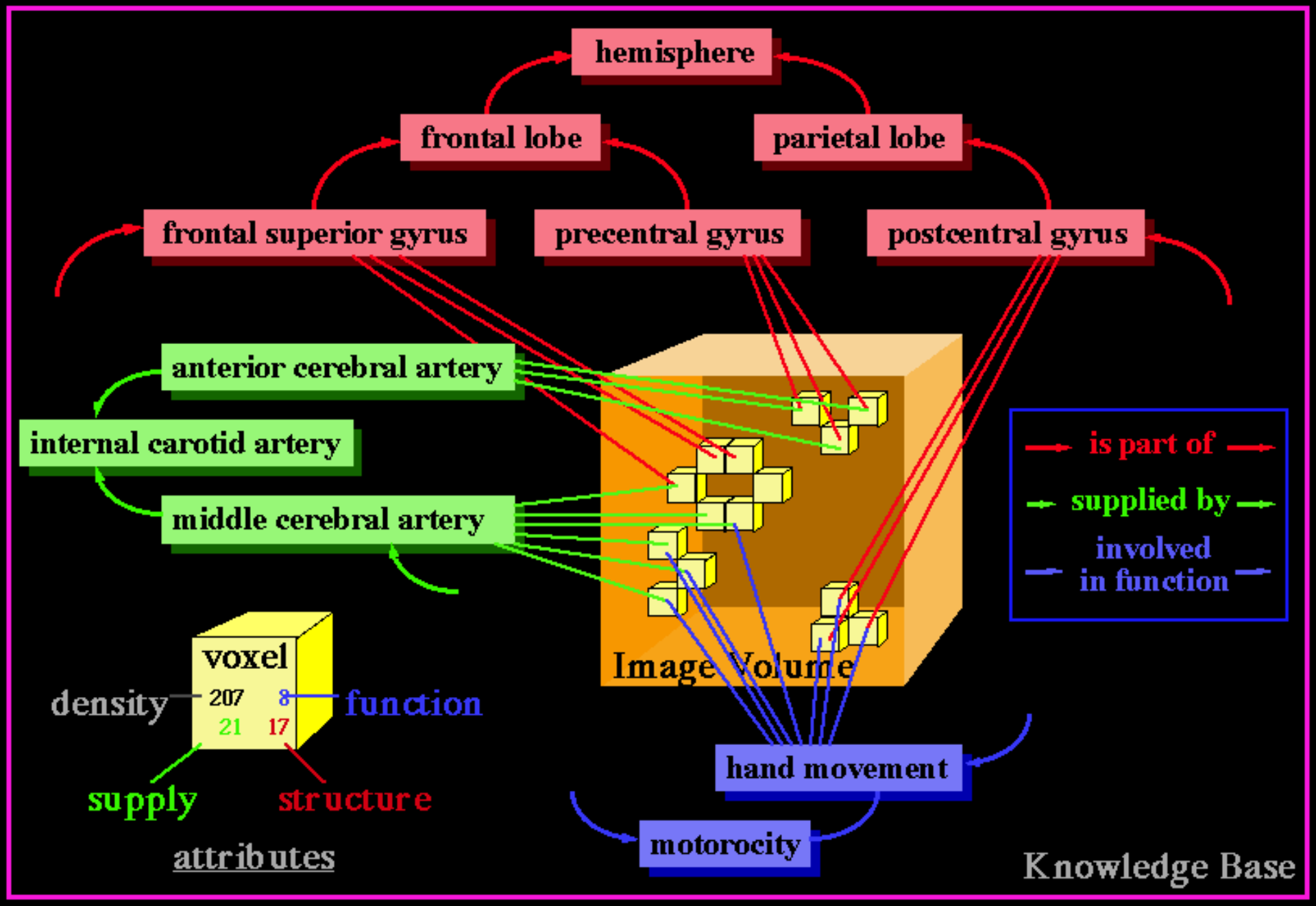
Individual Growth & Development



VoxelMan Electronic Anatomic Atlas



IMDM
Abteilung Informatik in der Medizin
Department of Computer Science in Medicine


hemisphere
frontal lobe
parietal lobe
frontal superior gyrus
precentral gyrus
postcentral gyrus
anterior cerebral artery
internal carotid artery
middle cerebral artery
Image Volume
is part of
supplied by
involved in function
voxel
density
207
8
function
21
17
supply
structure
attributes
hand movement
motorocity
Knowledge Base

Specificity

- Antithesis of variability
- Organism = genotype + epigenetics
- Genotype = DNA + others
- DNA = nuclear + mitochondrial
  - Genome is 99% the same for all
- Epigenetics = cytoplasm + environment
- At a macro level:
  - Anatomy + Physiology
  - Anatomy = species + individual variation + (growth & development) + degeneration + environment

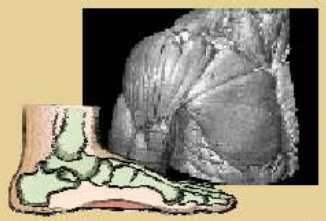
Medical technology

Degree of certainty

Telemedicine

Computer-Assisted Surgery

Receptor-Based Sensors



High-Speed, High-Resolution 3-D Imaging

Imaging Information Processing Systems

Telepresence Surgery

In-vivo Biosensors

Regenerative Implants

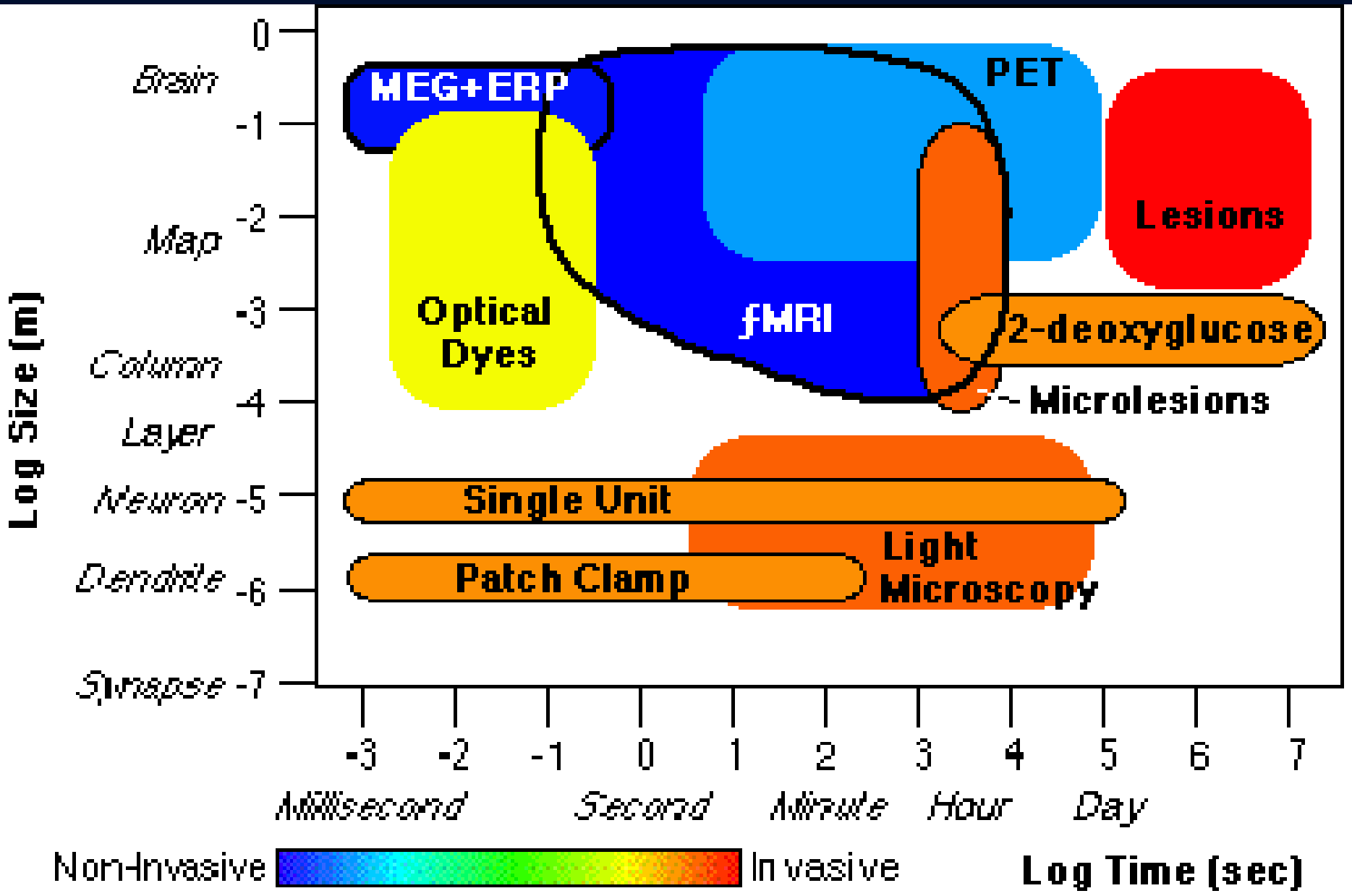
2000

2005

2010

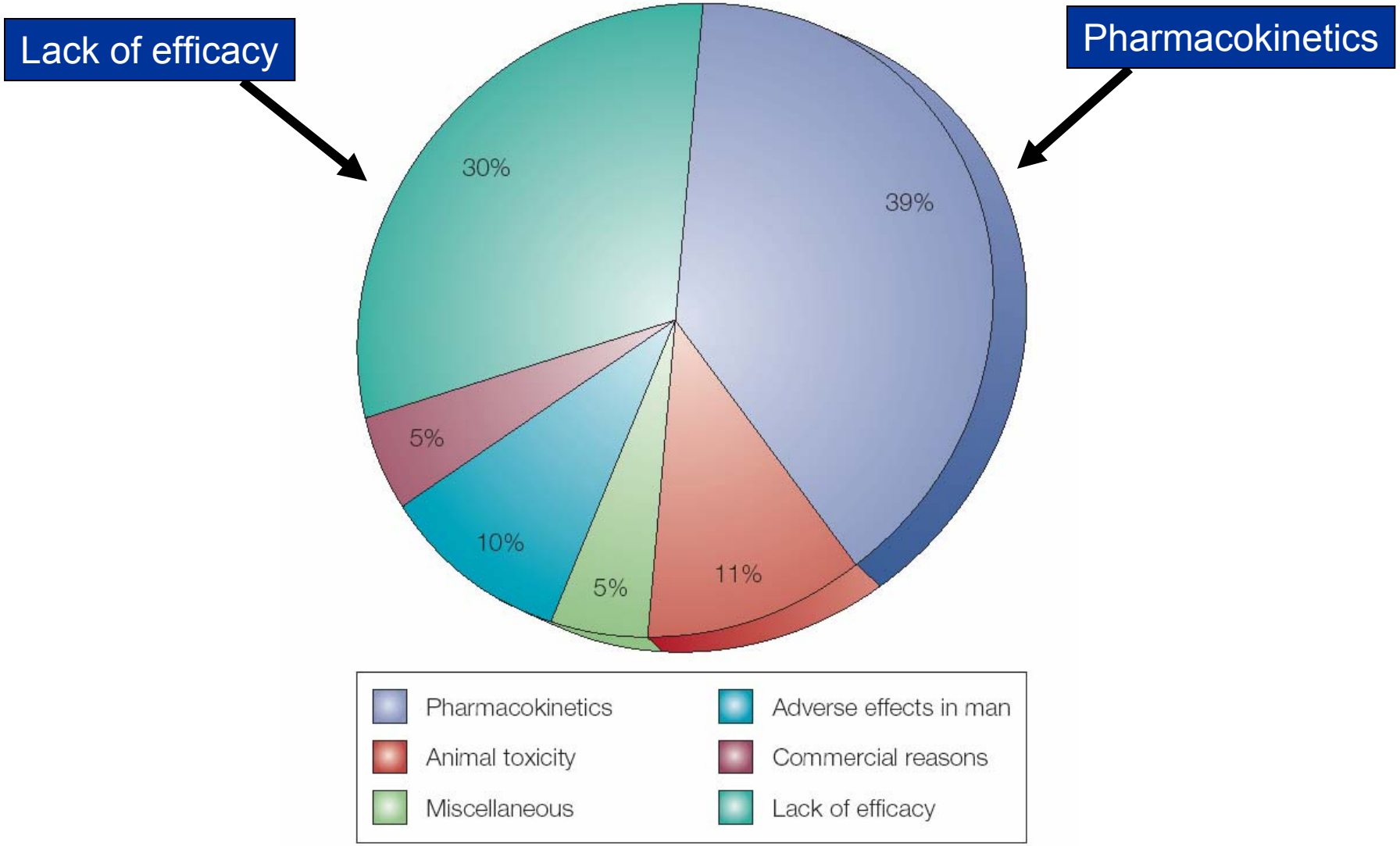
2020

© PricewaterhouseCoopers 2001



Churchland, P.S. and Sejnowski, T.J. "Perspectives on Cognitive Neuroscience." Science. 242(4879): 741-745, 1988.

Main reasons for attrition in drug development



Drug Development Pipeline

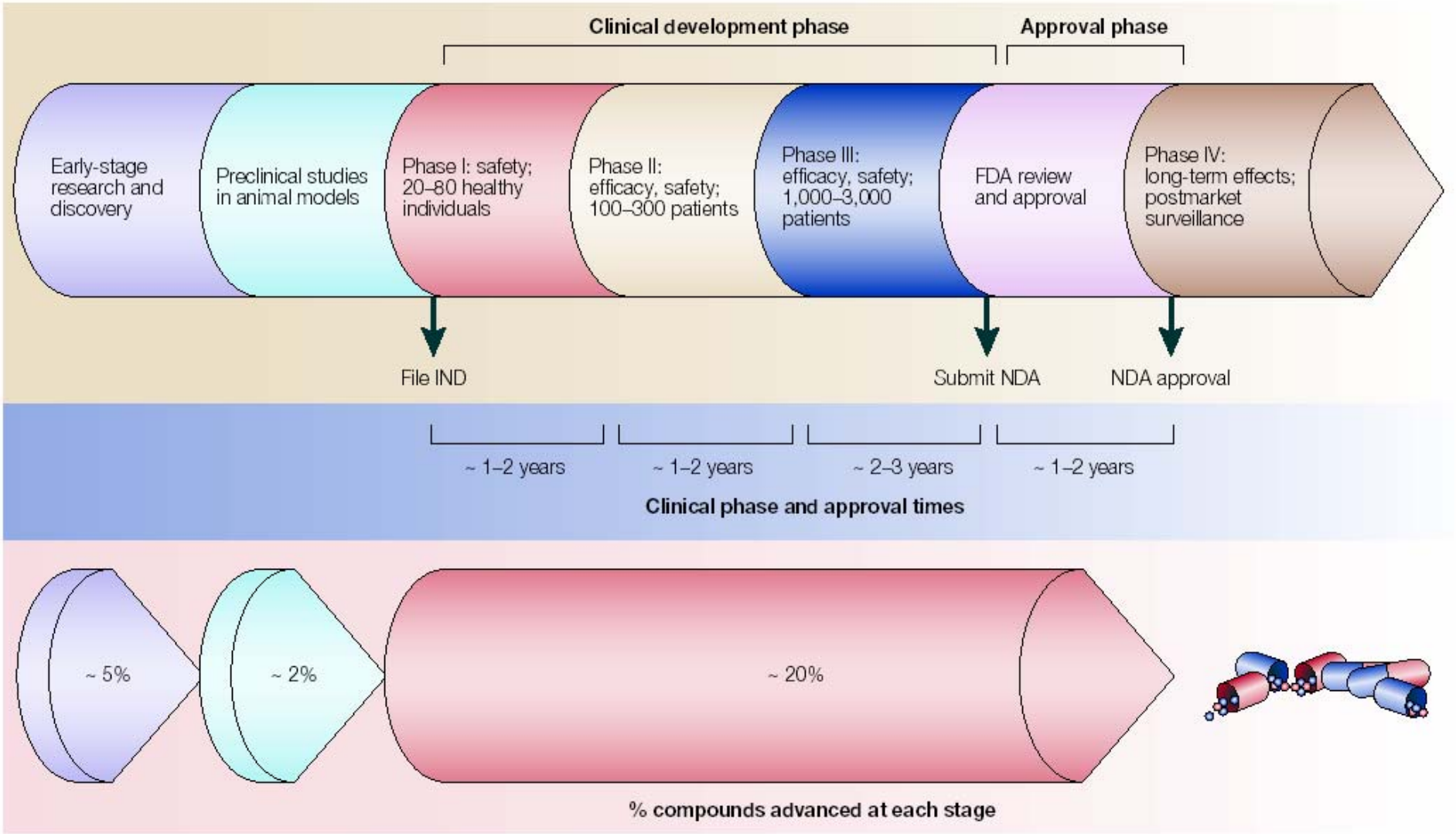
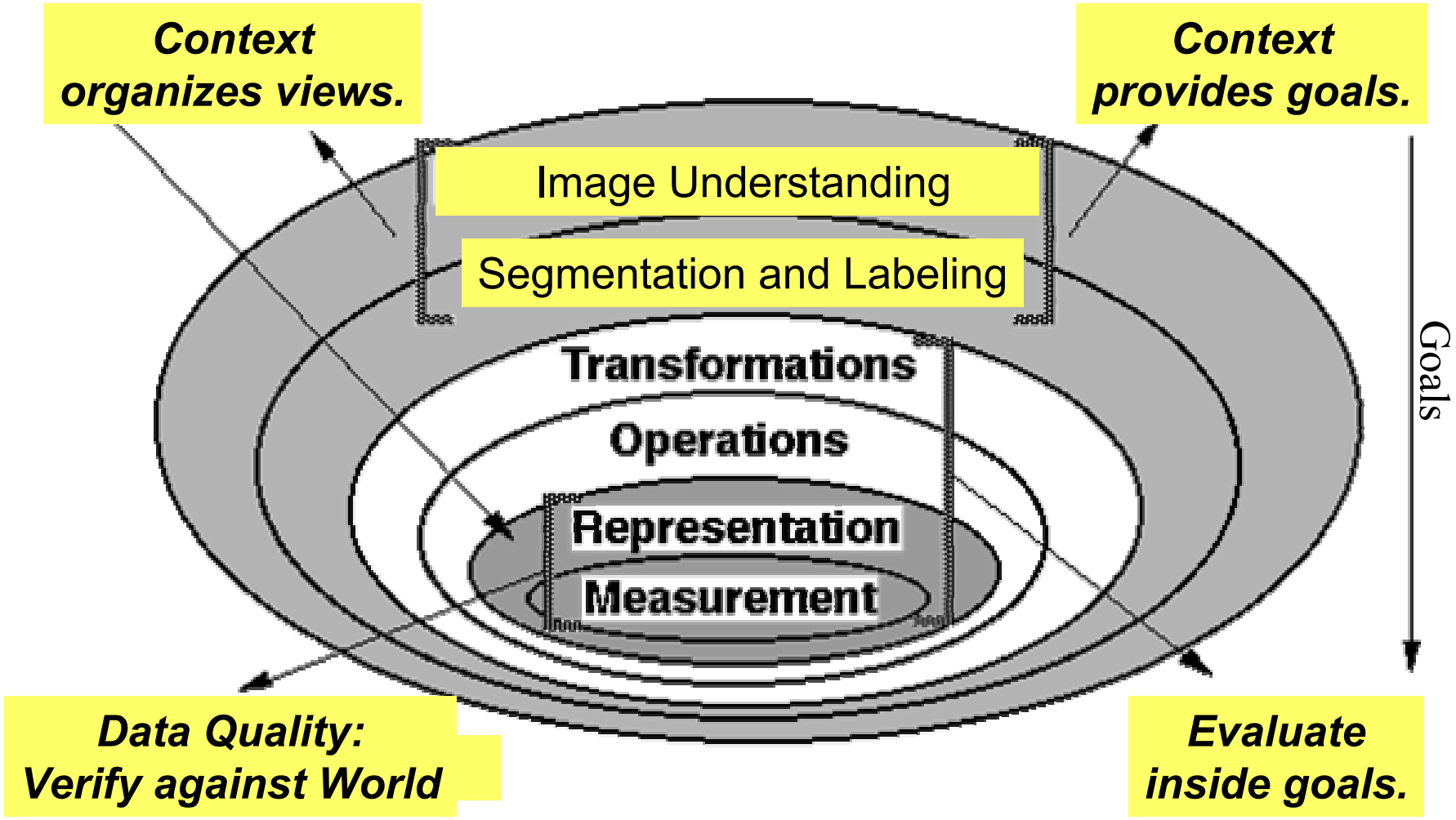


Figure 2 | **The drug development and approval chain**. FDA, US Food and Drug Administration; IND, Investigational New Drug Application; NDA, New Drug Application.

Questions that involve imaging

- What is the phenotype?
  - Size, shape
- How does this individual differ from others?
- What change has taken place?
  - Where is it?
  - How big is it?
- What is the target?
  - Magic bullet =
  - Access; specificity; toxicity; …
- What genes are expressed where?
  - Reconcile traits and genes
- How does the genetic makeup of individual affect function?
  - In the presence of a drug?
  - When subjected to an intervention?
  - What does this gene do?
- Given a specific gene, what variation does it explain?
  - What is the mechanism?
- Explain the functional consequences of an intervention
- …

Image Exploitation



Knowledge-based Decision Support Systems for Diagnosis and Therapy

Needs

- Integration
  - Multisensor
- Persistent infrastructure
- Security and confidentiality
- Speed
- Specificity
- Image exploitation

THE VIRTUAL PATIENT

Researchers hope that drug efficacy will be virtually assured in virtual clinical trials.

BY RANDALL C. WILLIS

IN pages past, this magazine has described the myriad wonders of the construction of virtual proteins based on gene and protein sequence alignments and the screening of virtual compounds against a database of drug targets. But as is so often the case in drug development, most of these virtual compounds fail to achieve their lofty goals when synthesized and exposed to the harshness of the real world and the complexity of the human body.

However, what if there were a way to test drugs in real-world scenarios while they were still in the virtual phase?

To answer this question, several groups of researchers have taken the real-world information that already exists and used it to develop virtual people on whom these compounds can be tested—biomedical Pinocchios, if you will. From these computational puppets, the researchers hope to predict the biological effects of the various test compounds in the hope of fine-tuning real-world assays and, ideally, eliminating dead-end leads from trials before they are even synthesized. To paraphrase the well-used quote, "Fail early and fail often."

VIRTUAL REALITY

To address this challenge, researchers have developed biological model systems that mimic the various healthy and diseased states against which the drug candidate is to be tested. In some cases, these models are simple cell- or tissue-based assays that examine how well a compound performs its assigned function. Often, however, animal models of a given disease state are developed by using gene knock-out experiments or by chemically or physically perturbing the animals' metabolism so that they have the human disease but from a different cause. But there are limits to this method.

"Pharma is all about model systems, whether it's comparative genomics, comparing zebrafish and *C. elegans* to humans, or using knock-out mice," says Michael French, vice president of Entelos (*www.entelos.com*), a company that specializes in computational modeling. "They're using a model system to understand a complex pathway network of the manipulation of a molecular target and the likely clinical outcome. But with an animal model, the problem is that the underlying problem is artificially induced. Simply put, rats don't get diabetes. So, although the pathways can be conserved across the species, the actual kinetic and quantitative aspects of them differ greatly."

This is where computational models of human physiology come into play. Researchers in academia and at companies like Entelos have developed various models, both diseased and healthy, that they hope will allow clinicians and scientists to test new drugs

FEBRUARY 2003 **MODERN DRUG DISCOVERY 35-40**

ADMET *IN SILICO* MODELLING: TOWARDS PREDICTION PARADISE?

Han van de Waterbeemd and Eric Gifford‡*

Following studies in the late 1990s that indicated that poor pharmacokinetics and toxicity were important causes of costly late-stage failures in drug development, it has become widely appreciated that these areas should be considered as early as possible in the drug discovery

Need early information on absorption, distribution, metabolism, excretion (ADME) and toxicity data (together called ADMETdata).

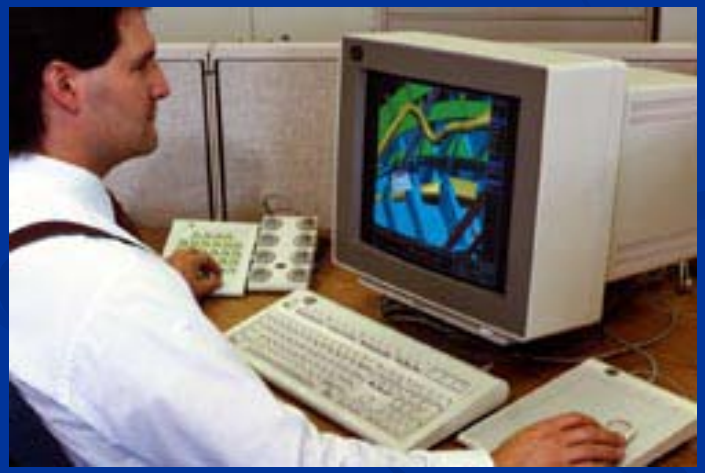
NATURE REVIEWS | DRUG DISCOVERY-VOLUME 2 | MARCH 2003 | **192-204**

Barriers

- Obstacles to collaboration
- Communications infrastructure
- Data sequestration
- Incompatible interfaces
- Isolation
- Tools

Boeing 777

- $5B project
- 10,000 engineers; 238 teams
- 5 years from concept to product
- 7,000 workstations
- 17 time zones
- > 1 M unique parts
- Used CATIA software



CATIA workstation

HBR
FROM THE HARVARD BUSINESS REVIEW
OnPoint

ARTICLE

Why do some of the best companies languish when markets change?

Because they insist on doing *only* what has worked in the past.

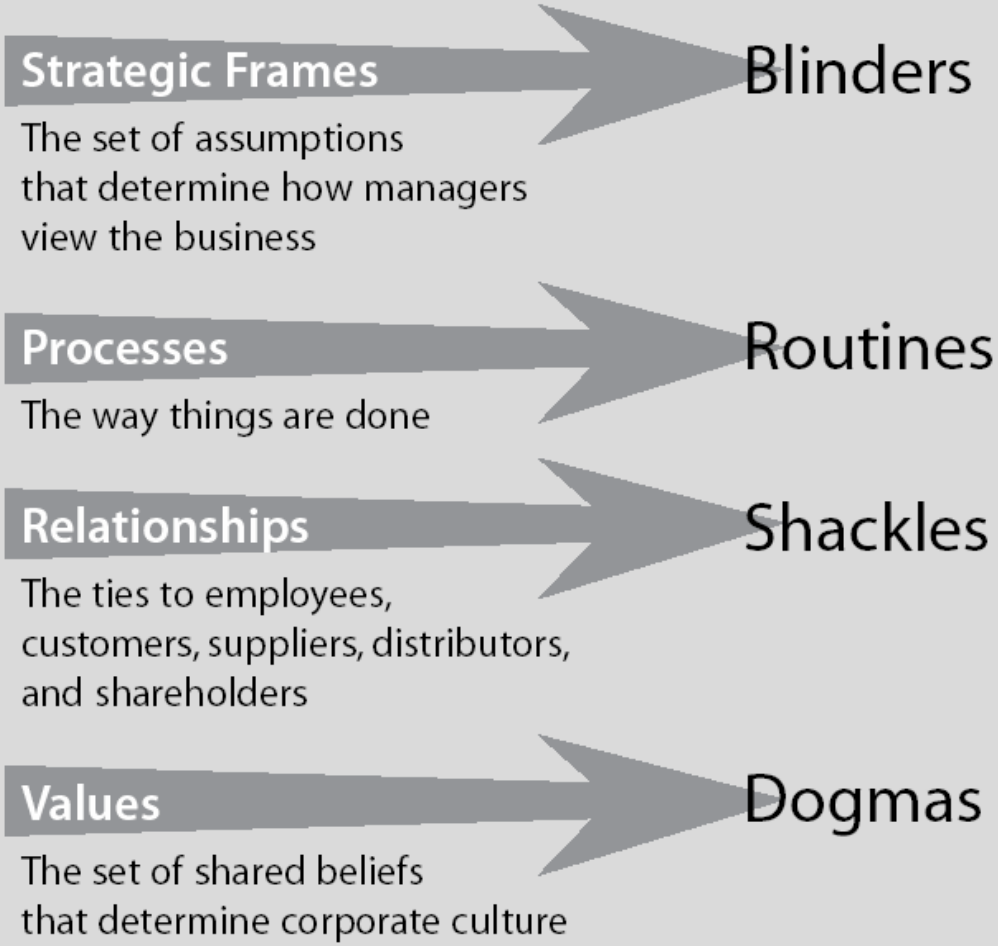
Why Good Companies Go Bad

by Donald N. Sull

Harvard Business Review July–August 1999

The Dynamic of Failure

Leading companies can become stuck in the modes of thinking and working that brought them their initial success. When business conditions change, their once-winning formulas instead bring failure.



Disastrous Effects of ACTIVE INERTIA

- **Strategic frames become blinders.**
  - Strategic frames shape how managers view their business; they help managers stay focused. But these frames can also blind managers to new options and opportunities.
- **Processes harden into routines.**
  - Established processes can become ends in themselves, even when they're no longer effective. People overlook better ways of working.
- **Relationships become shackles.**
  - Every company needs strong relationships with its constituencies—customers, suppliers, employees. When conditions change, however, these relationships can restrict flexibility.
- **Values harden into dogmas.**
  - A company's vibrant values unify and inspire its people. Over time, however, they can harden into rigid, self-defeating rules and regulations.

Life's Complexity Pyramid

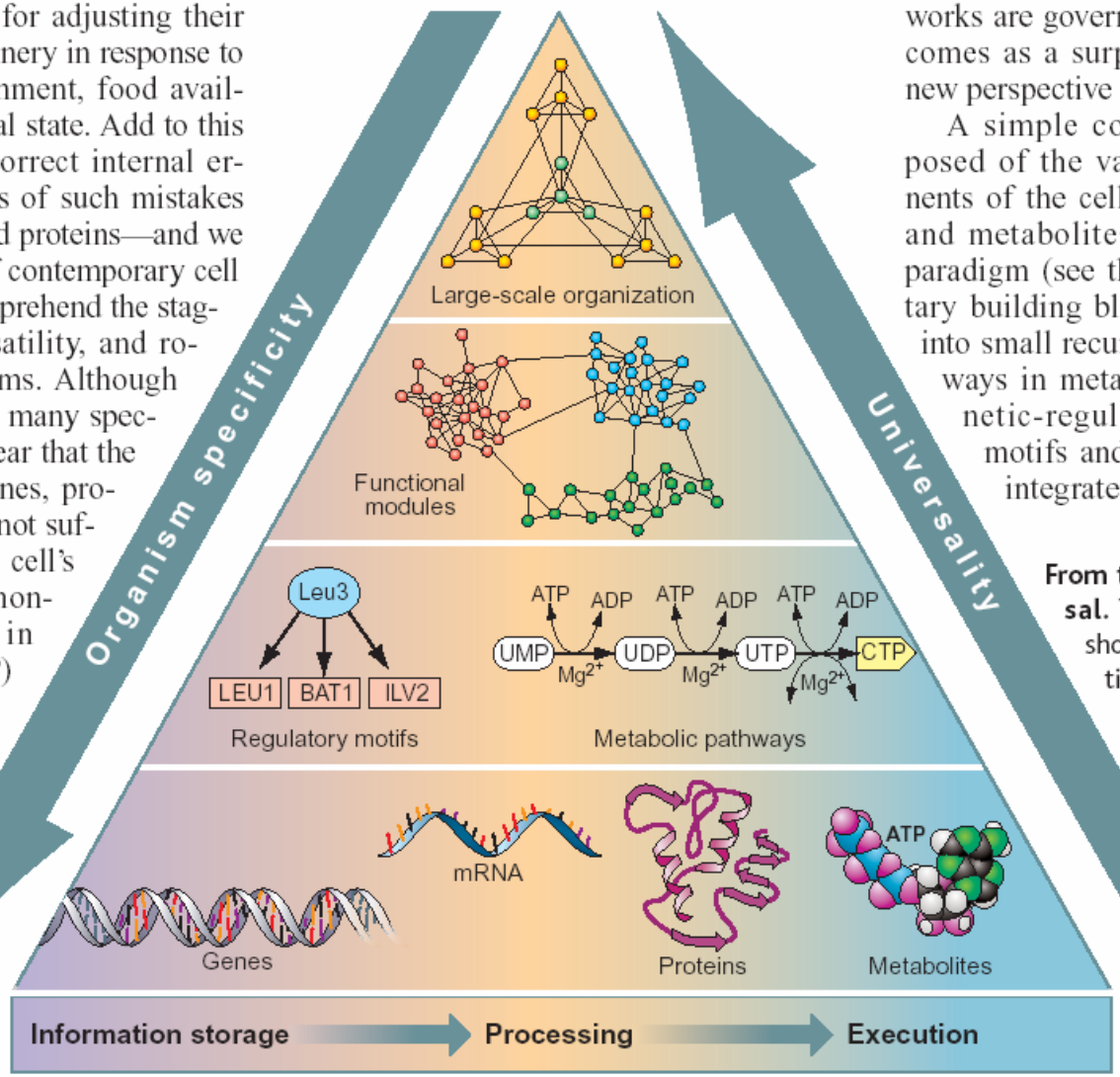
Zoltán N. Oltvai and Albert-László Barabási

Cells and microorganisms have an impressive capacity for adjusting their intracellular machinery in response to changes in their environment, food availability, and developmental state. Add to this an amazing ability to correct internal errors—battling the effects of such mistakes as mutations or misfolded proteins—and we arrive at a major issue of contemporary cell biology: our need to comprehend the staggering complexity, versatility, and robustness of living systems. Although molecular biology offers many spectacular successes, it is clear that the detailed inventory of genes, proteins, and metabolites is not sufficient to understand the cell's complexity (*1*). As demonstrated by two papers in this issue—Lee *et al.* (*2*) on page 799 and Milo *et al.* (*3*) on page 824—viewing the cell as a network of genes and proteins offers a viable strategy for addressing the complexity of living systems.

According to the basic dogma of molec-

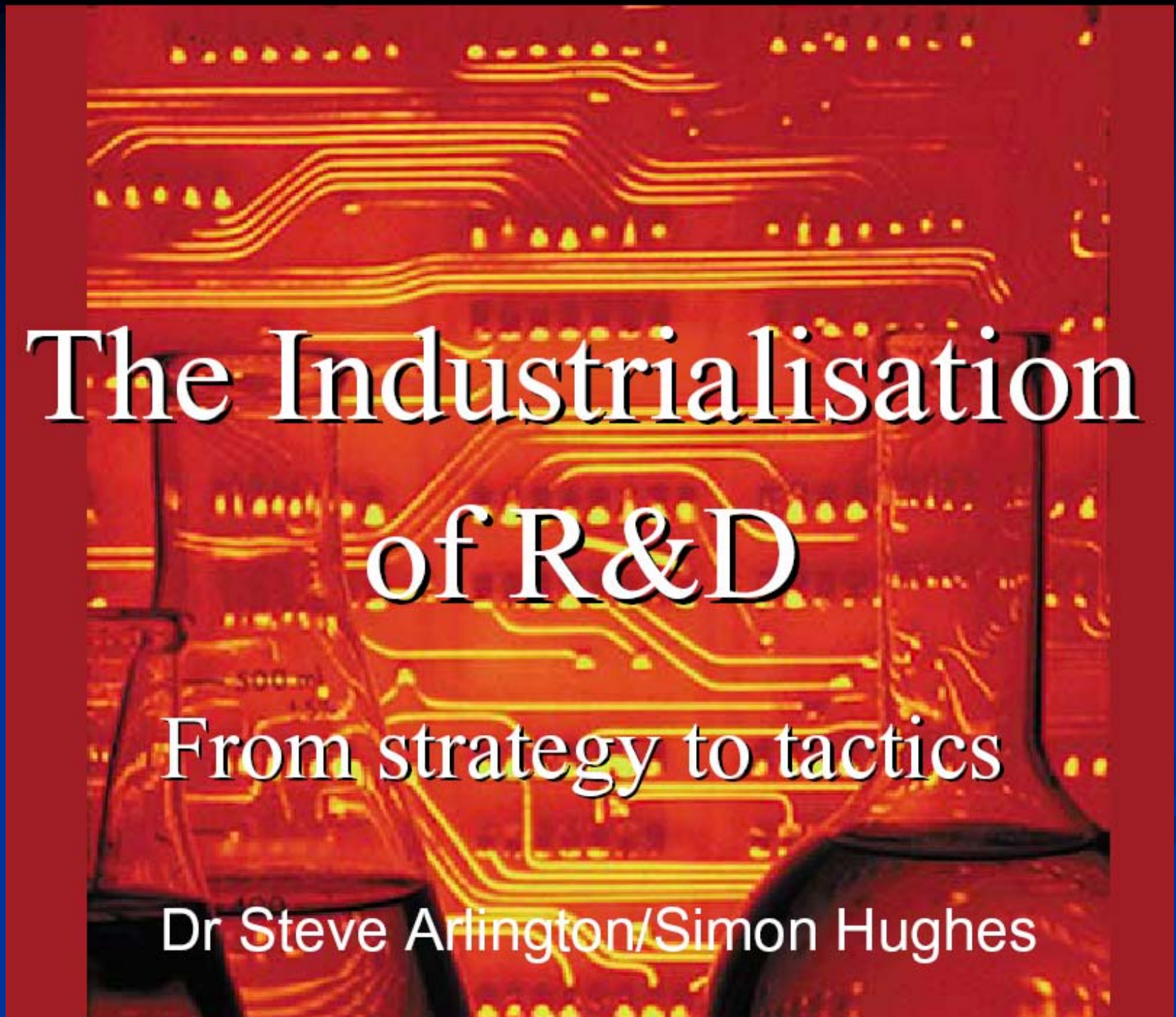
within large networks (*6*, *7*). There is clear evidence for the existence of such cellular networks: For example, the proteome organizes itself into a protein interaction network and metabolites are interconverted through an intricate metabolic web (*7*). The finding that the structures of these networks are governed by the same principles comes as a surprise, however, offering a new perspective on cellular organization.

A simple complexity pyramid composed of the various molecular components of the cell—genes, RNAs, proteins, and metabolites—summarizes this new paradigm (see the figure). These elementary building blocks organize themselves into small recurrent patterns, called pathways in metabolism and motifs in genetic-regulatory networks. In turn, motifs and pathways are seamlessly integrated to form functional mod-

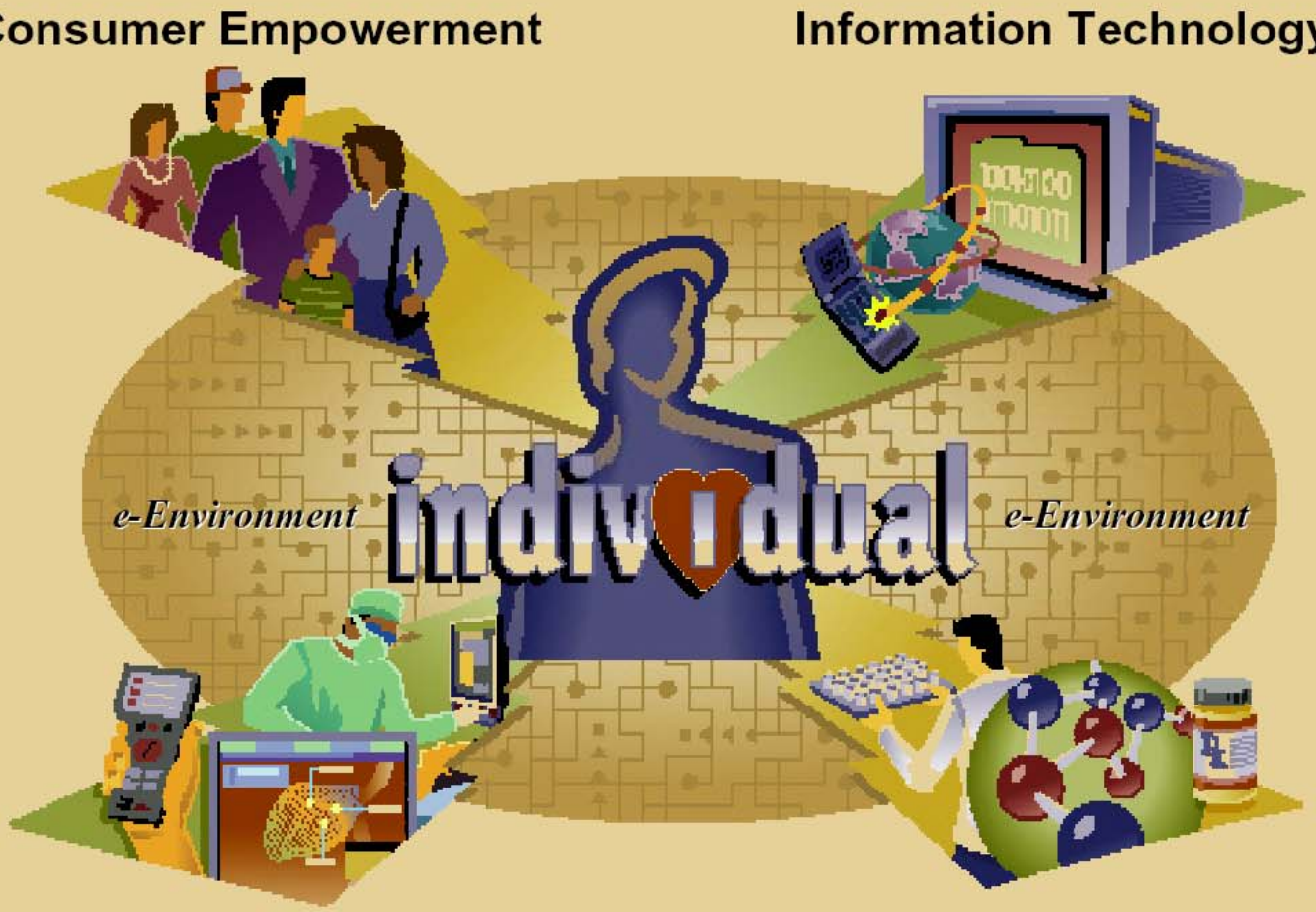


From the particular to the universal. The bottom of the pyramid shows the traditional representation of the cell's functional organization: genome, transcriptome, proteome, and metabolome (level 1). There is remarkable integration of the various layers both at the regulatory and the structural level. Insights into the logic of cellular organization can be achieved when we view

Science, Vol 298, 25 Oct 02; 762-3

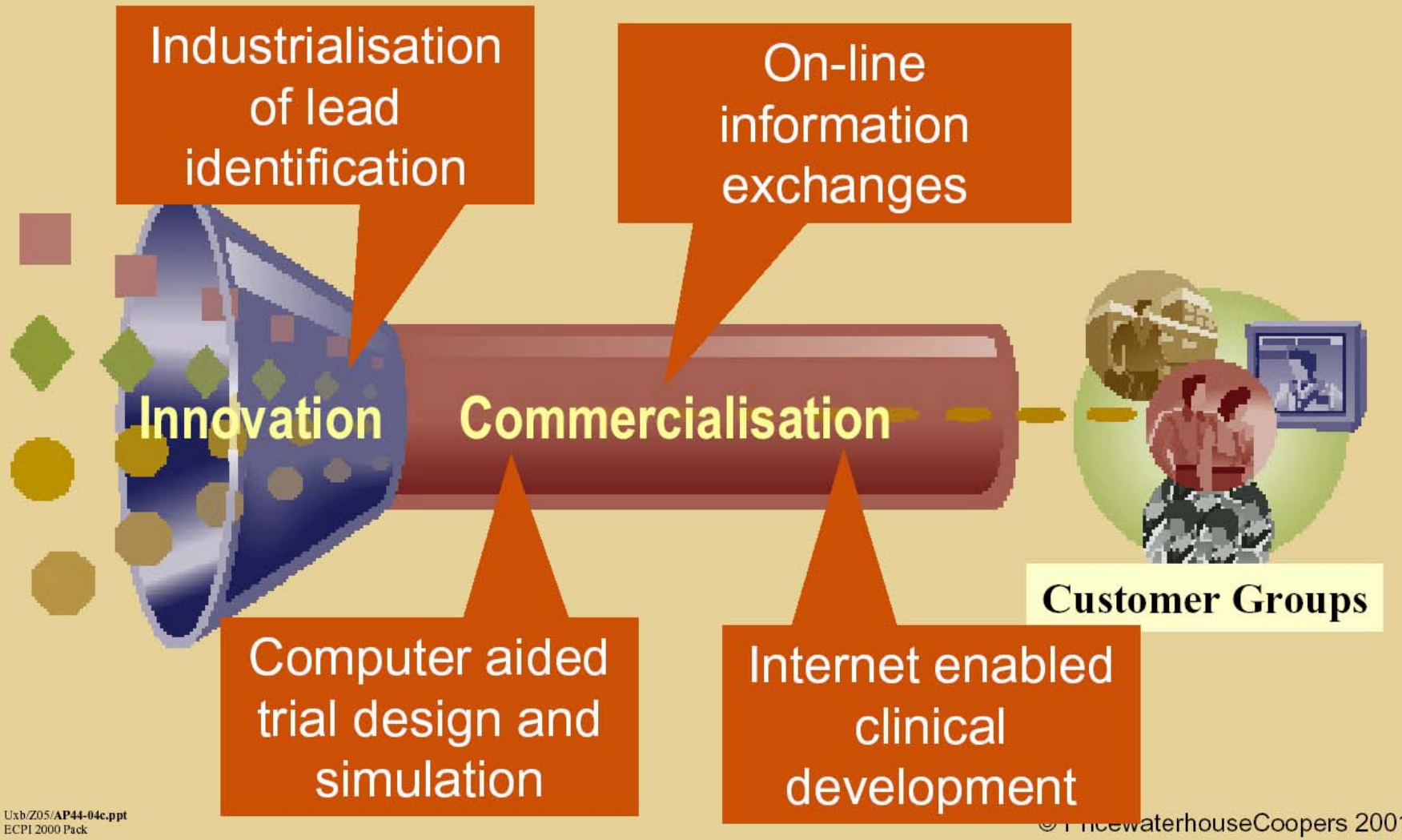

The Industrialisation of R&D
From strategy to tactics
Dr Steve Arlington/Simon Hughes

Forces for change



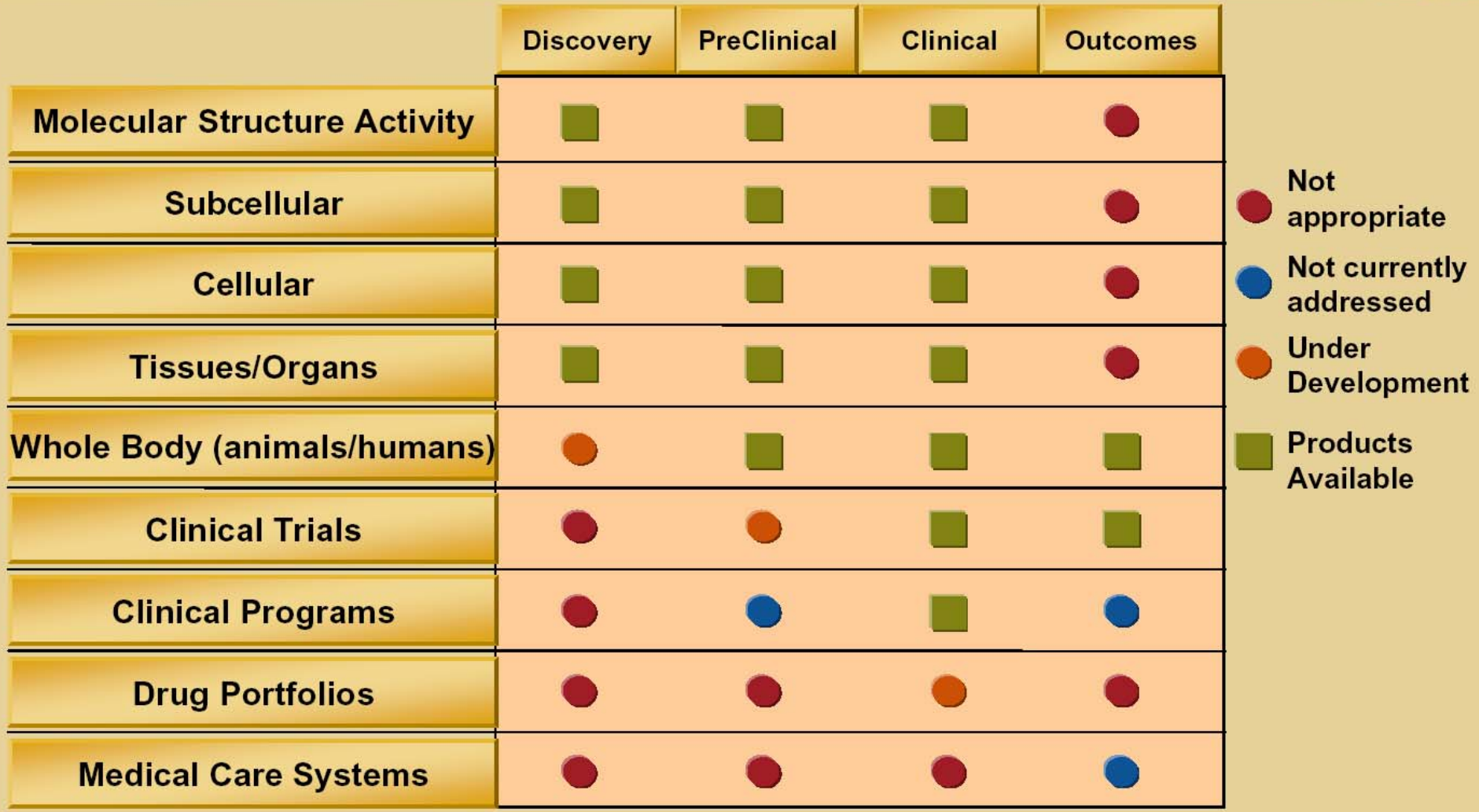
© PricewaterhouseCoopers 2001

4 key areas for change over next 3 years



© PricewaterhouseCoopers 2001

Simulation - a rapidly emerging technology



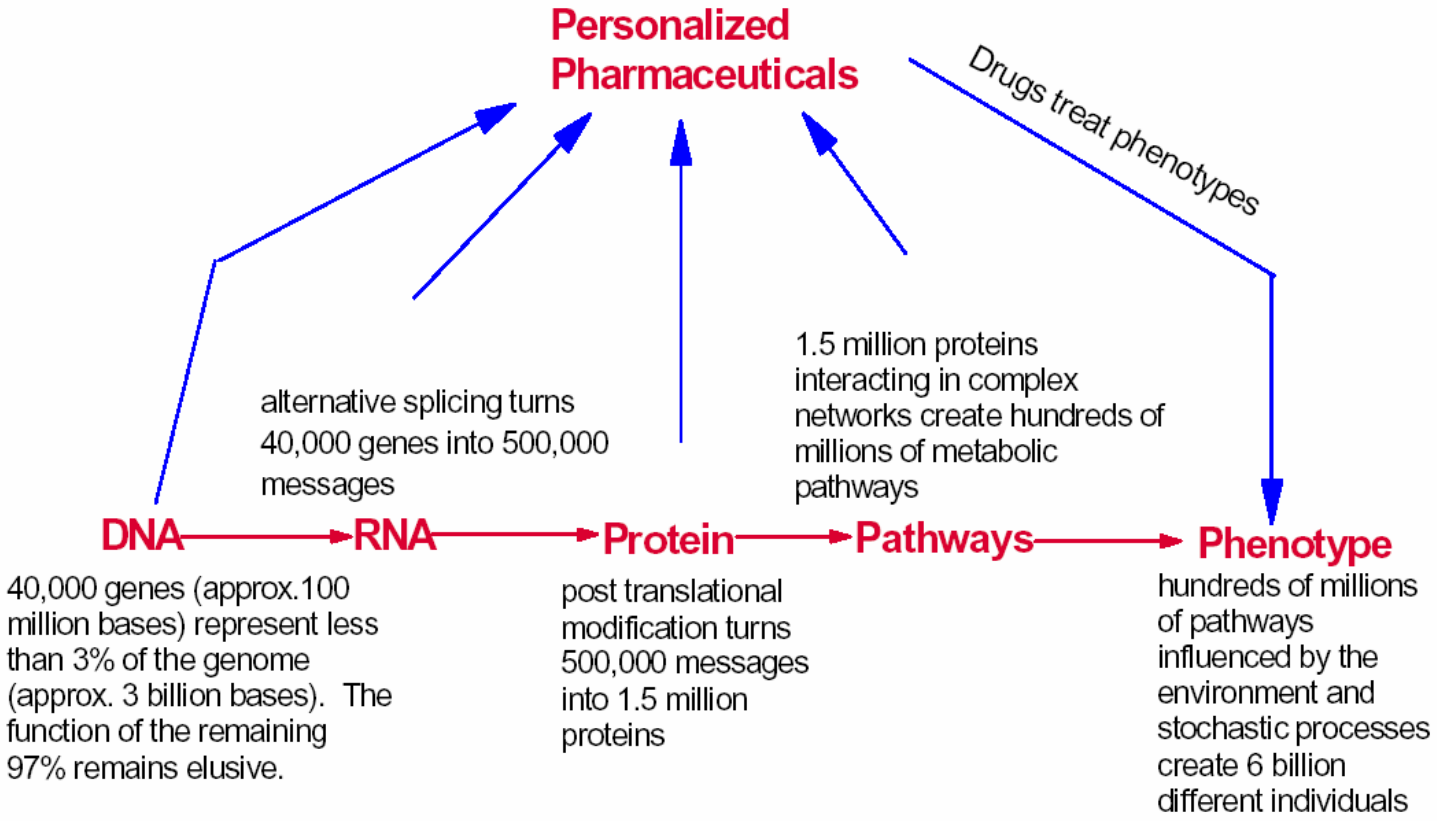
© PricewaterhouseCoopers 2001

Metrics

- Signal
- Noise
  - Random
  - Artifacts
  - Systematic
- Diagnostic performance
- Variation
  - Short term = test-retest reproducibility
  - Long term = stability
- Most important is “Value Creation”

New Sources of Value Creation

Rational drug development requires managing enormous complexity. Pharmaceutical companies are beginning to differentiate themselves on the power of their information technology platforms. IT Platform intellectual property is likely to be more valuable than content (gene sequences, metabolic pathways, protein structures, etc.)



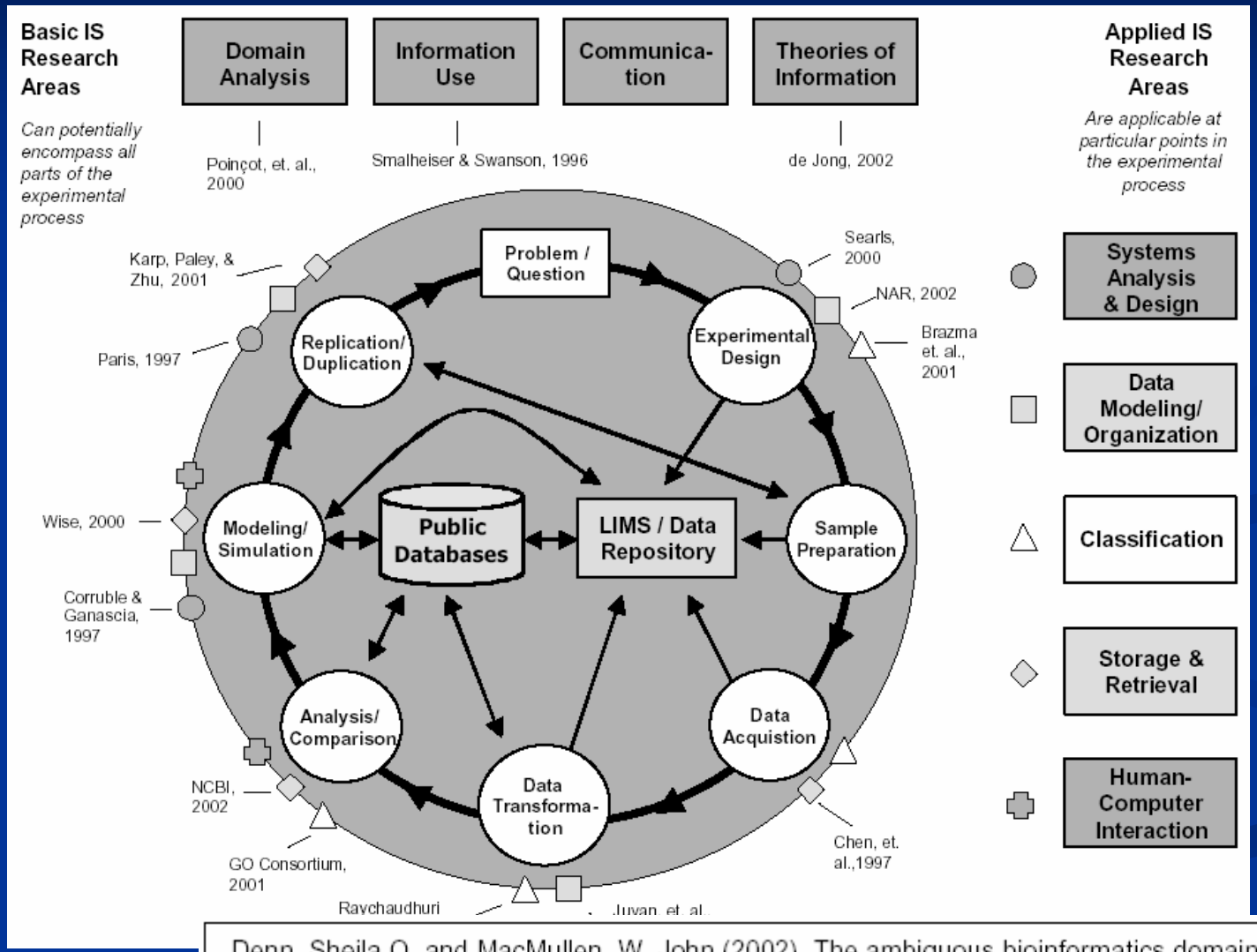
Historically, 220 targets have generated $3trillion of value. Industrialized genome sequencing has created a target rich, lead poor environment that will slowly reverse over the next several years as in-silico biology drives the discovery of new lead compounds.

IBM Life Sciences

What is a Biological Database?

- A biological database is
  - a large, organized body of persistent data,
  - usually associated with software to
    - update,
    - query, and
    - retrieve the data stored within the system.

Molecular Biology Experimental Cycle and Information Science

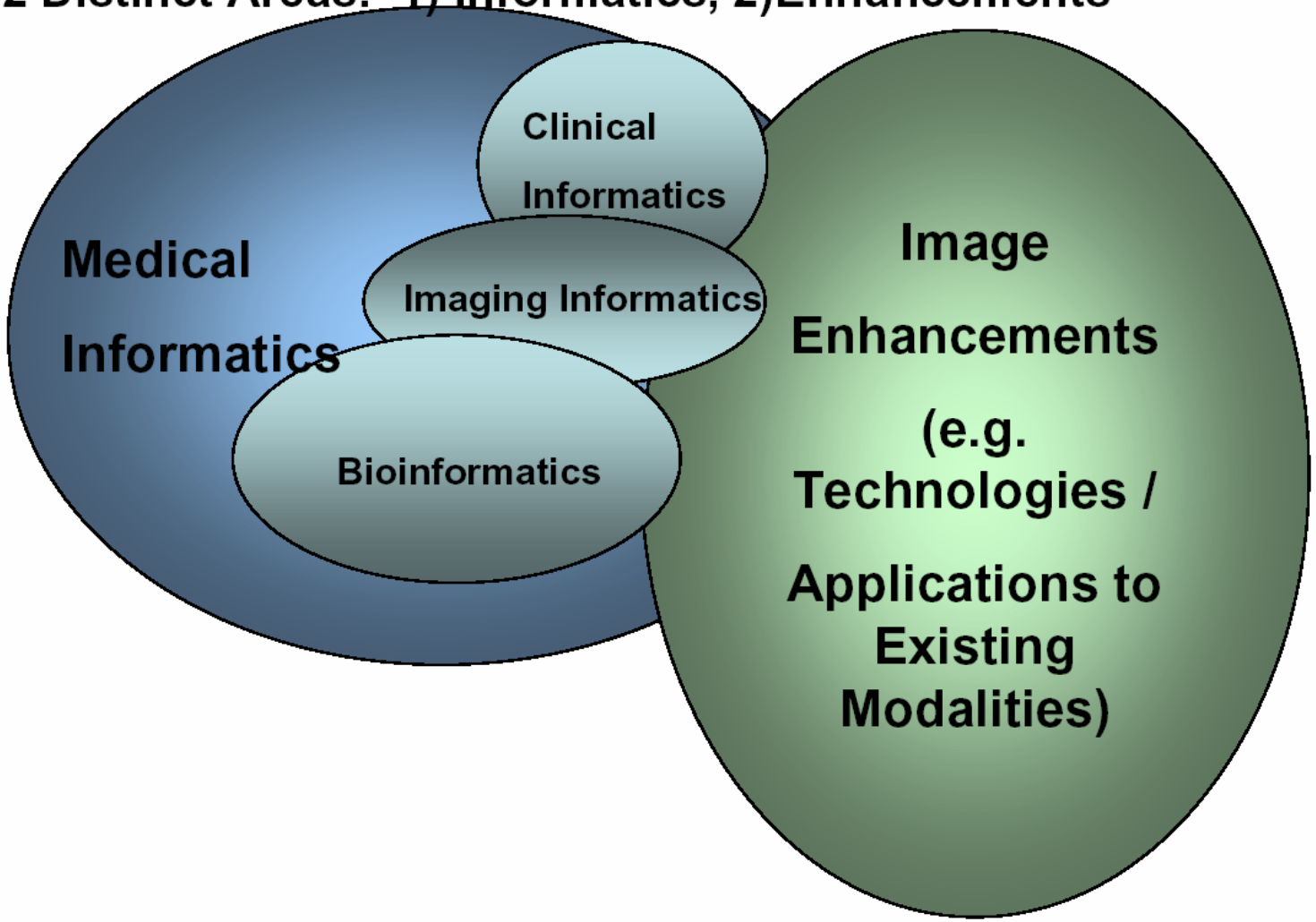


Denn, Sheila O. and MacMullen, W. John (2002). The ambiguous bioinformatics domain: A conceptual map of information science applications to molecular biology. In *Proceedings of the 65th Annual Meeting of the American Society for Information Science & Technology (ASIS&T)*, pp. 556-558

Medical Imaging

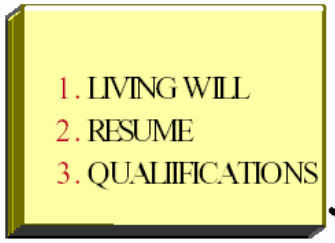
How Do We Define Medical Imaging?

2 Distinct Areas: 1) Informatics, 2)Enhancements

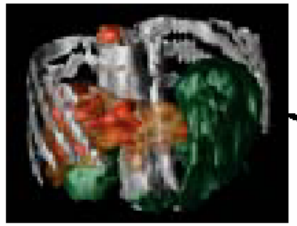


IBM
Life
Sciences

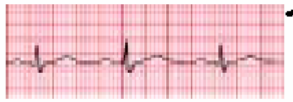
Clinical Records Will Be Complex Heterogeneous Objects



Digital Images



Hospital transaction events



Integrated Data

Expression Arrays (various tisues)

Clinical Record

Analysis lab notes

"What protocols were used for tumors which prouduced similar staining sections, and were from same age group as my patient, and with same "Yakamura" polymorphism in her genes?"

IBM Life Sciences

What we can do now

- Macroimaging – 2D and 3D
- Modest temporal variation – milliseconds to seconds
- Churchland - Sejnowski map
- CT; MR; US; PET/SPECT; …
  - Single modalities for in vivo humans
- Some animal imaging
- A few contrast agents
- Small changes in abundant molecules
- Guide therapy with images

What we cannot do now

- Apply and integrate multiple modalities (multisensor fusion)
- Localize molecules within the cell (in vivo)
- Integrate image and non-image data
- Verify much of published work with available data and tools
- Control the specificity of imaging tests

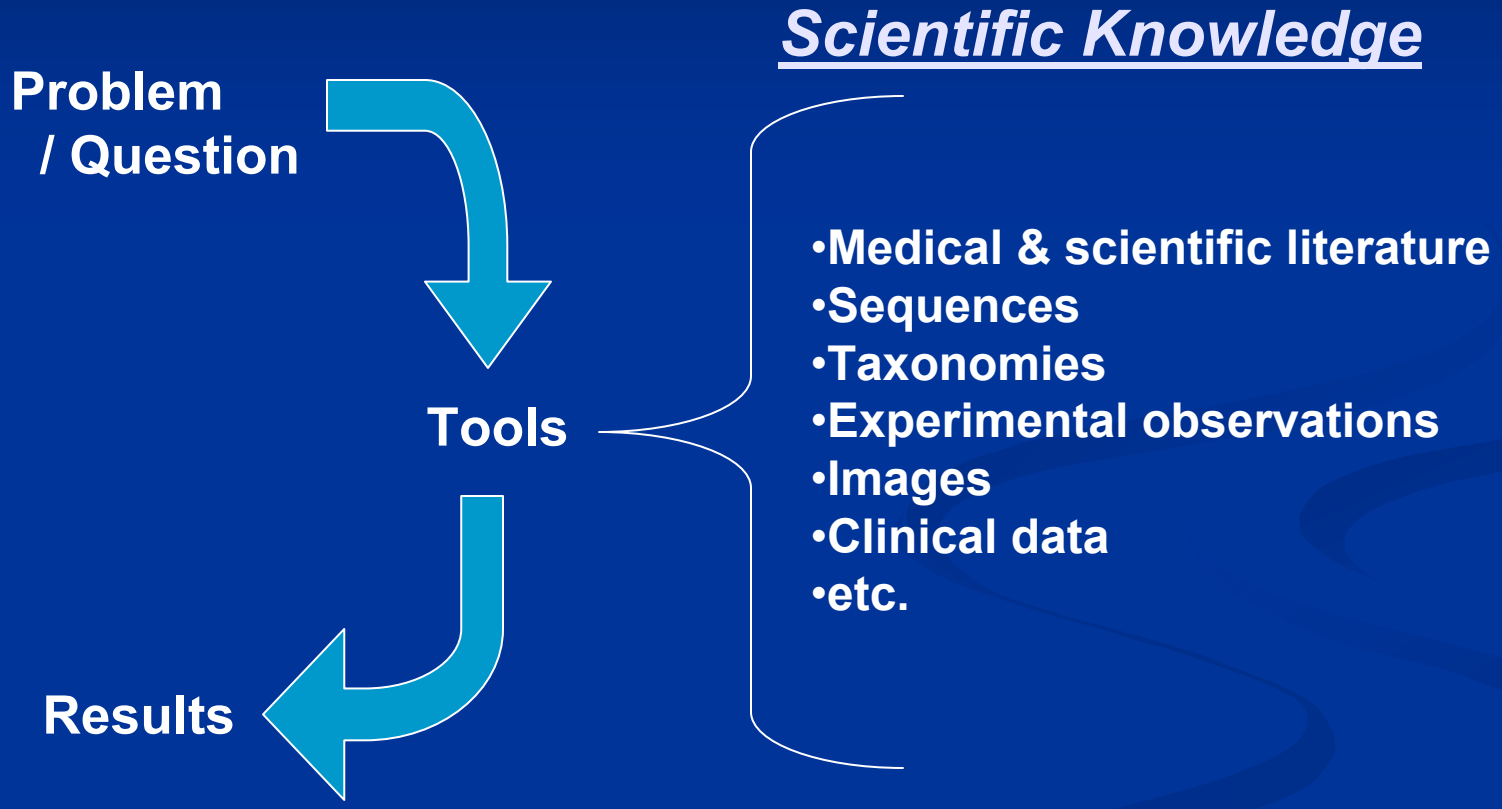
What we could do to improve the situation

- Integrate = KISS
- Persistent infrastructure
- Guide technology developments with specific biological questions
- Develop specifications and design systems that are task specific
  - Optimize and validate systems according to task

Areas of concern

- Data acquisition capabilities increase faster than infrastructure to organize and use the information we gather
- Disconnection between in vivo images and "mainstream" biological knowledge resources (e.g., genome and text databases, among others)
- Biomedical imaging science appears to lag behind neuroscience and genomics/proteomics in the integrated information infrastructure

Investigation



The New England Journal of Medicine

CYCLIN E AND SURVIVAL IN PATIENTS WITH BREAST CANCER

KHANDAN KEYOMARSI, PH.D., SUSAN L. TUCKER, PH.D., THOMAS A. BUCHHOLZ, M.D., MATTHEW CALLISTER, M.D., YE DING, PH.D., GABRIEL N. HORTOBAGYI, M.D., ISABELLE BEDROSIAN, M.D., CHRISTOPHER KNICKERBOCKER, M.S., WENDY TOYOFUKU, B.S., MICHAEL LOWE, B.S., THADDEUS W. HERLICZEK, M.D., AND SARAH S. BACUS, PH.D.

ABSTRACT

Background Cyclin E, a regulator of the cell cycle, affects the behavior of breast-cancer cells. We investigated whether levels of cyclin E in the tumor correlated with survival among patients with breast cancer.

Methods Tumor tissue from 395 patients with breast cancer was assayed for cyclin E, cyclin D1, cyclin D3, and the HER-2/*neu* oncogene with the use of Western blot analysis. Full-length, low-molecular-weight, and total cyclin E were measured. Immunohistochemical assessments of cyclin E were also made of 256 tumors. We sought correlations between levels of these molecular markers and disease-specific and overall survival.

Results The median follow-up was 6.4 years. A high level of the low-molecular-weight isoforms of cyclin E,

THE prognosis in patients with newly diagnosed breast cancer is determined primarily by the presence or absence of metastases in draining axillary lymph nodes.[1] However, in approximately one third of women with breast cancer who have negative lymph nodes, the disease recurs, and about one third of patients with positive lymph nodes are free of recurrence 10 years after local–regional therapy.[2,3] These data highlight the need for more sensitive and specific prognostic indicators.

A number of biologic factors have been used to refine risk categories in breast cancer. We have focused on the role of cyclin E in determining the virulence and metastatic potential of tumor cells.[4-8] In normal dividing cells, cyclin E regulates the transition from

N Engl J Med, Vol. 347, No. 20, **1566-1575** November 14, 2002

overall survival. Total cyclin E levels and the level of low-molecular-weight forms of cyclin E as measured by Western blotting but not by immunohistochemical analysis proved to be strongly associated with survival among patients with breast cancer.

METHODS

Tissue Samples and Study Patients

Tumor tissue was obtained from a centralized reference laboratory (Quantitative Diagnostic Laboratories). A total of 430 samples consisting of a minimum of 100 mg of breast-cancer tissue were available. Each patient had received a diagnosis of breast cancer between 1990 and 1995 at 1 of 12 hospitals in the Chicago area. Specimens were shipped to the Wadsworth Center research laboratories for Western blot analysis. This study was approved by the institutional review board of the Wadsworth Center.

The reference laboratory also provided base-line pathological and demographic data (obtained from the individual hospitals), as well as the steroid-receptor status, the DNA index, and the proliferation index (as described below). Information concerning clinical staging and survival was obtained from the tumor registries of each hospital. Patients whose death was clearly documented to be due to breast cancer were considered to have died of breast cancer; other deaths were considered not to have been caused by breast cancer. The data presented here are from 395 patients for whom data on outcome were available.

Hormone-Receptor, DNA, and Proliferation Assays

The procedures for the hormone-receptor and proliferation assays

values obtained from normal tissues. To
pression was scored as high if the value w
than the highest value for normal breast
normal tissue samples were examined.
weight and total cyclin E, specimens wi
were classified as high. All normal-cell
cyclin D3 and HER-2/*neu*. On Western
ues for these proteins clustered into thre
as negative, low level, or high level. Den
were used to standardize for equal protei
ples assayed. The Western blot analysis an
tioned biologic markers were performed
unaware of the patients' outcomes.

Immunohistochemical Studies

A subgroup of 256 samples of tumo
immunohistochemical analysis with the
polyclonal antibody to cyclin E.[5] We us
corresponding to amino acids 381 to 4
tigen in the affinity purification. This po
the same epitope as monoclonal HE12 a
Western blots to detect both the full-l
weight isoforms of cyclin E.[5,7,8] Snap-froz
in Optimal Cutting Temperature compo
tervals, placed on coated slides, fixed, and
viously described.[22-24] At least two represe
each patient with breast cancer were ex
from 0 to 10 on the basis of the intensi
centage of tumor cells stained. In 15 case
were tested along with tumor tissue. Sc
controls ranged from 0 to 2. The tumo
ignated as having eithe

RAND

www.rand.org/publications/MR/MR954

Handbook of Human Tissue Sources: A National Resource of Human Tissue Samples

Elisa Eiseman
Susanne B. Haga

Reviewer's comments

Contents

Appendix A

QUANTITY OF STORED TISSUE SAMPLES IN THE UNITED STATES

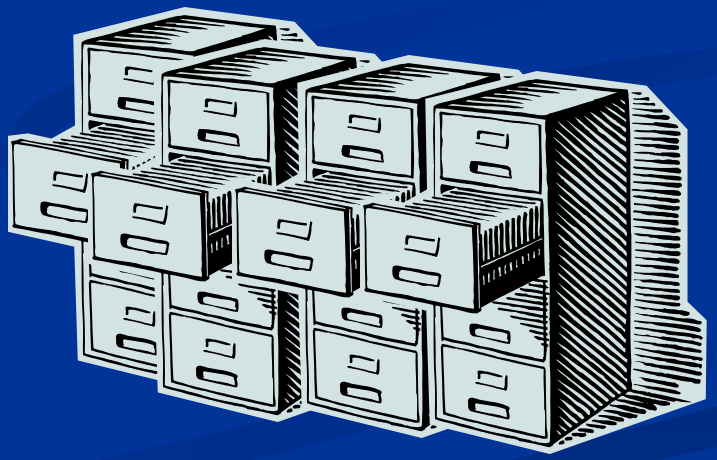
Type of Repository/Institution	Number of Cases	Number of Specimens	Cases/Year
Large Tissue Banks, Repositories, and Core Facilities			
AFIP DNA Specimen Repository		>2.8 million	10,000
AFIP National Pathology Repository	>2.5 million	>92 million	50,000
Brain and Tissue Banks for Developmental Disorders	2,507	34,943	
Cancer Tissue Bank—VAMC Minneapolis		>2,000	
Newborn Screening Laboratories			
50 states, District of Columbia, Puerto Rico, and Virgin Islands	>13.5 million	>13.5 million	<10,000 to >500,000
Forensic DNA Banks			
48 states with Forensic DNA Banks	1.4 million	1.4 million	

Image Repositories

- Open access to image archives is rare
- Image archives are not organized for research queries
- Image archives are not linked to other forms of biological data
- In general, there is no equivalent of a "specimen repository" for images
- This is a major problem for imaging research

Silo of Data

=



The Problem

- Most biological knowledge is stored in databases
- Creation, expansion, and integration of these databases has become central to the advancement of biology and medicine
- Many image databases are isolated "silos"
- Medical imaging is unique in that there are few publicly accessible databases, links to mainstream biological knowledge collections are absent, and there are few (software) tools available that allow you to use them

(protein/nucleic acid) Sequence Data and Molecular Biology Journals

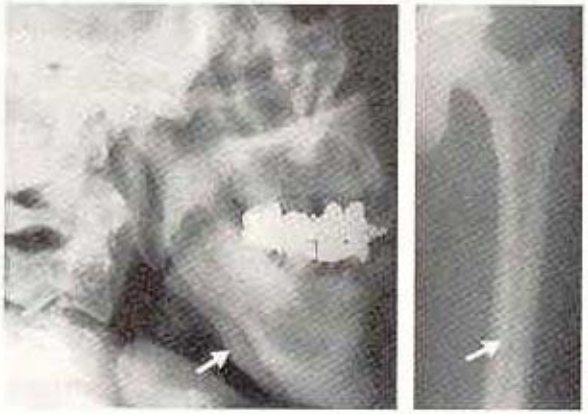
- Prior to publication, peer-reviewed molecular biology journals require that the authors deposit their data sets in a publicly-accessible archive and obtain an Accession Number.
- The Accession Number is included with the publication (both printed and electronic form)
- In many cases, the software tools used to analyze the sequence data are in the public domain

Why are imaging databases important?

- Images contain the phenotype
- In other fields (e.g., astronomy, geoscience, neuroscience, …), the integration of image (and other) databases has had a revolutionary effect
  - Coalescence of the scientific community
  - Open the field to rapid technological advancement
  - Possible to address questions that could not otherwise be answered (e.g., trans-species, multiscale, ad hoc group collaboration)

This Week in the Journal

May 16, 2002



High Bone Density Due to an *LRP5* Mutation

Osteoporosis can be caused by a loss-of-function mutation in the gene for low-density lipoprotein receptor–related protein 5 (*LRP5*). In this study, the authors, reasoning that a gain-of-function mutation in the same gene might be associated with high bone density, performed biochemical and genetic analyses of a kindred with high bone density, a prominent mandible, and torus palatinus. Genetic analysis revealed an *LRP5* mutation, the substitution of valine for glycine at codon 171, that segregated with the clinical findings. In vitro studies demonstrated that the defect in LRP5 resulted in changes in signaling events with other molecules that normally interact with this receptor-related protein, resulting in increased bone density.

The findings suggest that molecules that interact with LRP5 *may provide targets for the treatment of osteoporosis.*

The New England Journal of Medicine

Copyright © 2002 by the Massachusetts Medical Society

VOLUME 346 MAY 16, 2002 NUMBER 20

HIGH BONE DENSITY DUE TO A MUTATION IN LDL-RECEPTOR–RELATED PROTEIN 5

LYNN M. BOYDEN, PH.D., JUNHAO MAO, PH.D., JOSEPH BELSKY, M.D., LYLE MITZNER, M.D., ANITA FARHI, R.N., MARY A. MITNICK, PH.D., DIANQING WU, PH.D., KARL INSOGNA, M.D., AND RICHARD P. LIFTON, M.D., PH.D.

ABSTRACT

Background Osteoporosis is a major public health problem of largely unknown cause. Loss-of-function mutations in the gene for low-density lipoprotein receptor–related protein 5 (*LRP5*), which acts in the Wnt signaling pathway, have been shown to cause osteoporosis–pseudoglioma.

Methods We performed genetic and biochemical analyses of a kindred with an autosomal dominant syndrome characterized by high bone density, a wide

OSTEOPOROSIS is a major public health problem, and its prevalence is increasing.[1-3] In the United States, nearly 1 million fractures occur annually in people over the age of 65 years, the majority of which are due to osteoporosis.[1,4] Osteoporotic fractures are associated with substantial morbidity, and the estimated rate of death in the first year after a hip fracture is 25 to 30 percent.[5,6]

Bone mass, a major determinant of the risk of os-

es an autosomal dominant disorder characterized by high bone density, torus palatinus, and a wide, deep mandible.

Our in vitro and in vivo studies show that the $LRP5_{V171}$ mutation increases Wnt signaling. The mutation impairs antagonism of Wnt signaling by Dkk-1 in vitro, and the levels of fibronectin, a downstream target of Wnt signaling, are increased in vivo in patients with this mutation. These findings indicate that unopposed Wnt signaling due to loss of action of a

gle propeller of the low-density lipoprotein (LDL) receptor in humans, mice, rats, pigs, hamsters, and rabbits. Moreover, glycine is also found at this position in the first propeller of the *Drosophila melanogaster* LDL-receptor–related protein homologue, *arrow*. In addition, glycine is present at this position in a wide range of other YWTD propellers, including those in other LDL-receptor–related proteins, as well as those in the epidermal growth factor precursor, the very-low-density lipoprotein receptor, and the vitellogenin receptor in fruit flies and mosquitos (protein sequences are available at http://www.ncbi.nlm.nih.gov/entrez). The evolutionary conservation of this glycine residue is strong evidence of the functional importance of its mutation in our kindred.

Protein sequences are available at

ENTREZ

Molecular Studies

If this mutation indeed causes gain of LRP5 function and increased Wnt signaling, downstream target genes in the Wnt signaling pathway should show increased expression in vivo. A direct transcriptional target of Wnt signaling is the extracellular matrix protein fibronectin.[31] Fibronectin levels were markedly elevated in the affected members of our kindred, with

It is striking that the same mutation is associated with nonsyndromic high bone mass in one family and syndromic high bone mass in the other. These findings suggest that alleles of other genes or environmental factors influence phenotypic manifestations of the mutation and that other phenotypes in kindreds with autosomal dominant high bone mass may also arise from

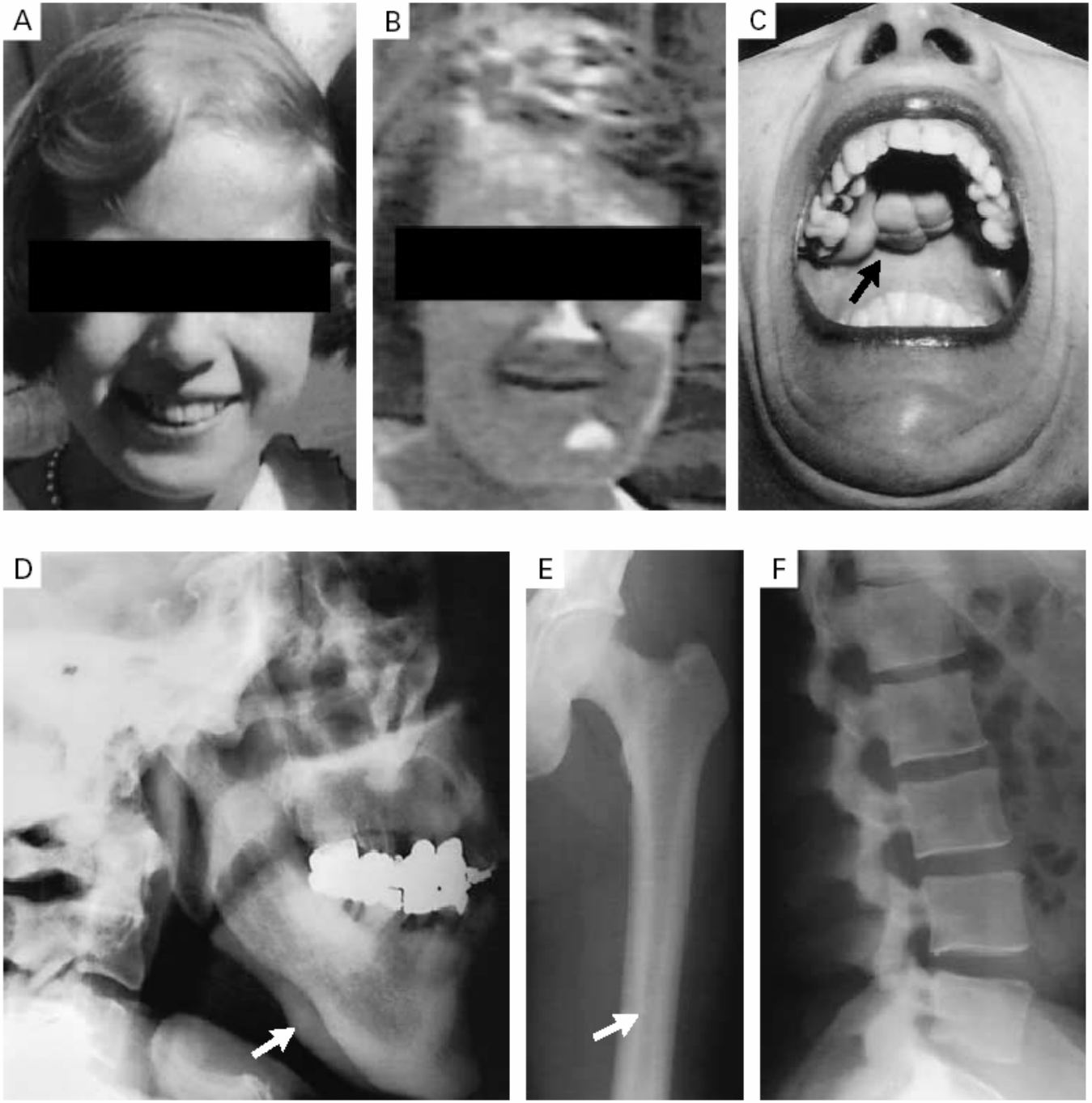
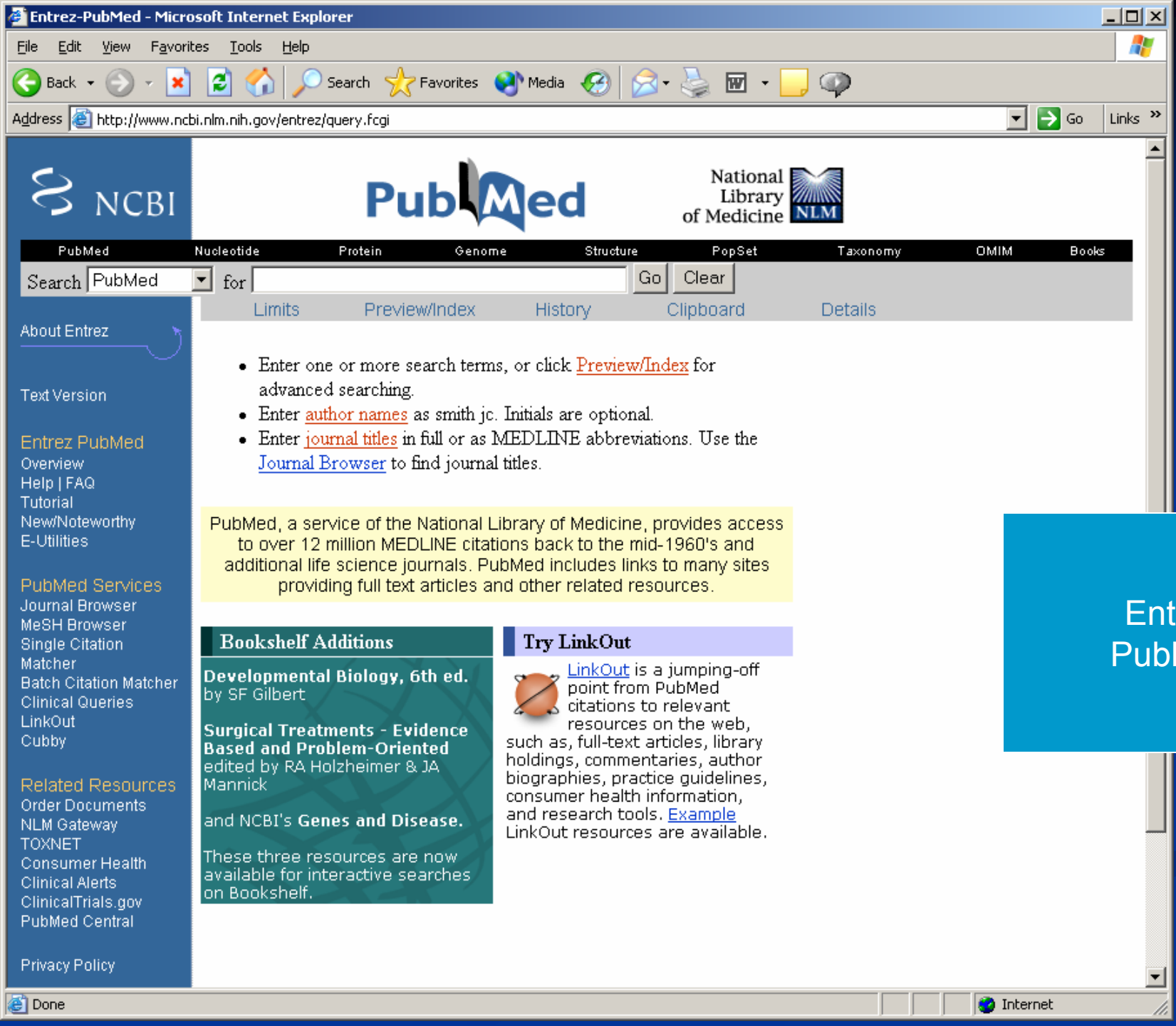


Figure 1. Clinical and Radiographic Features of Affected Members of the Kindred.

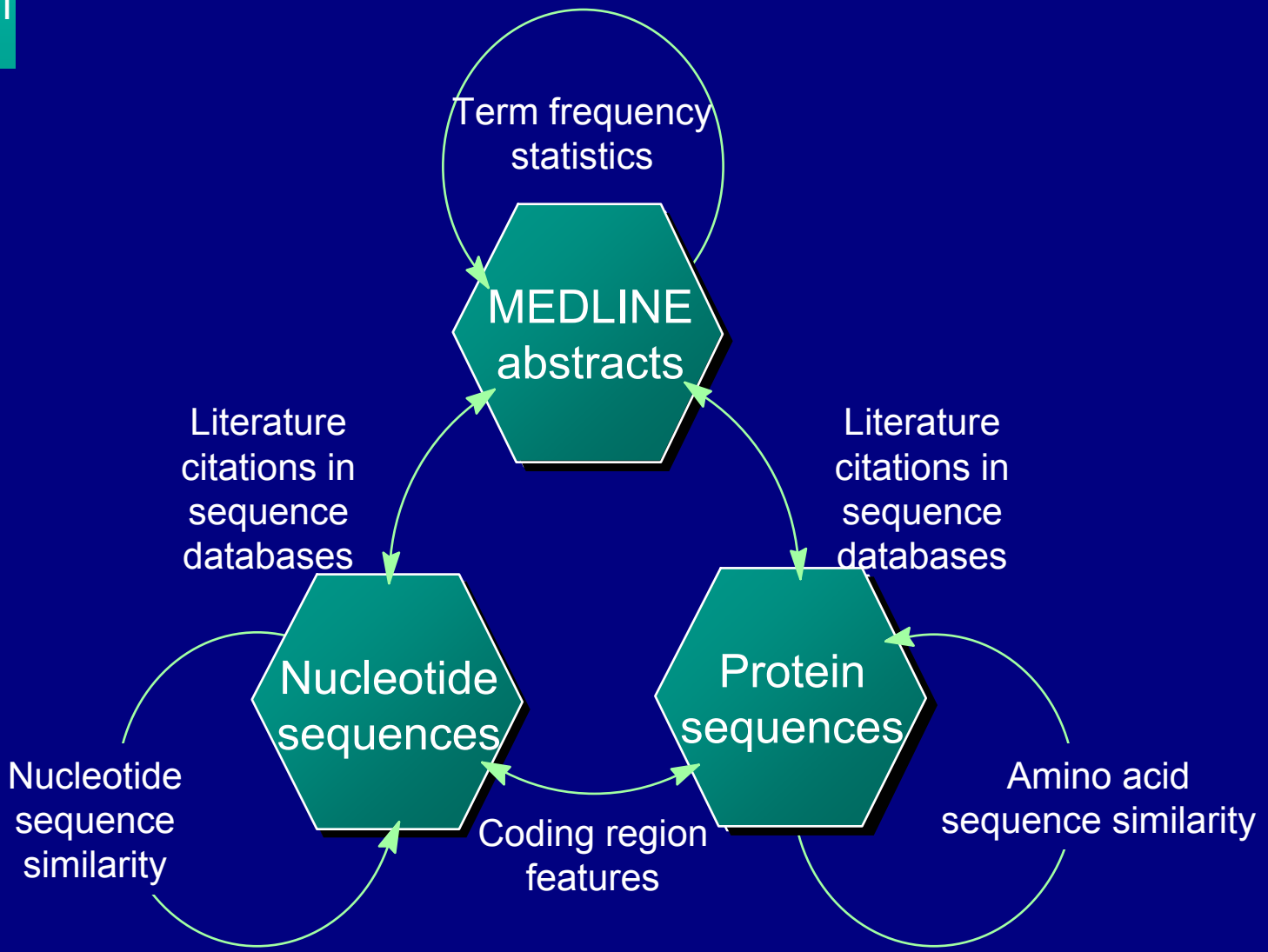
Figure 1

Page 1515



Entrez PubMed

Entrez: Pathway to Discovery



Entrez Increases Discovery Space

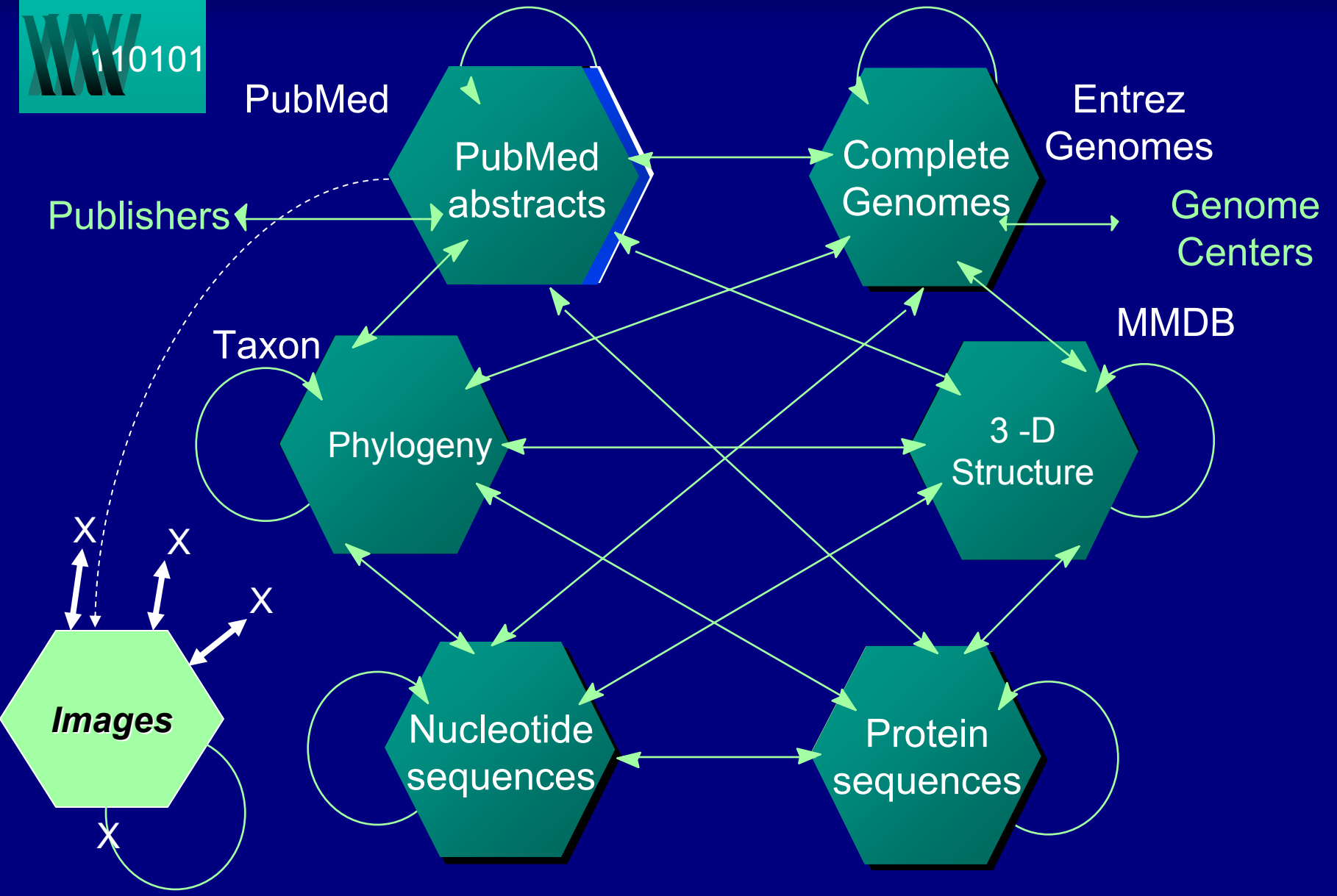
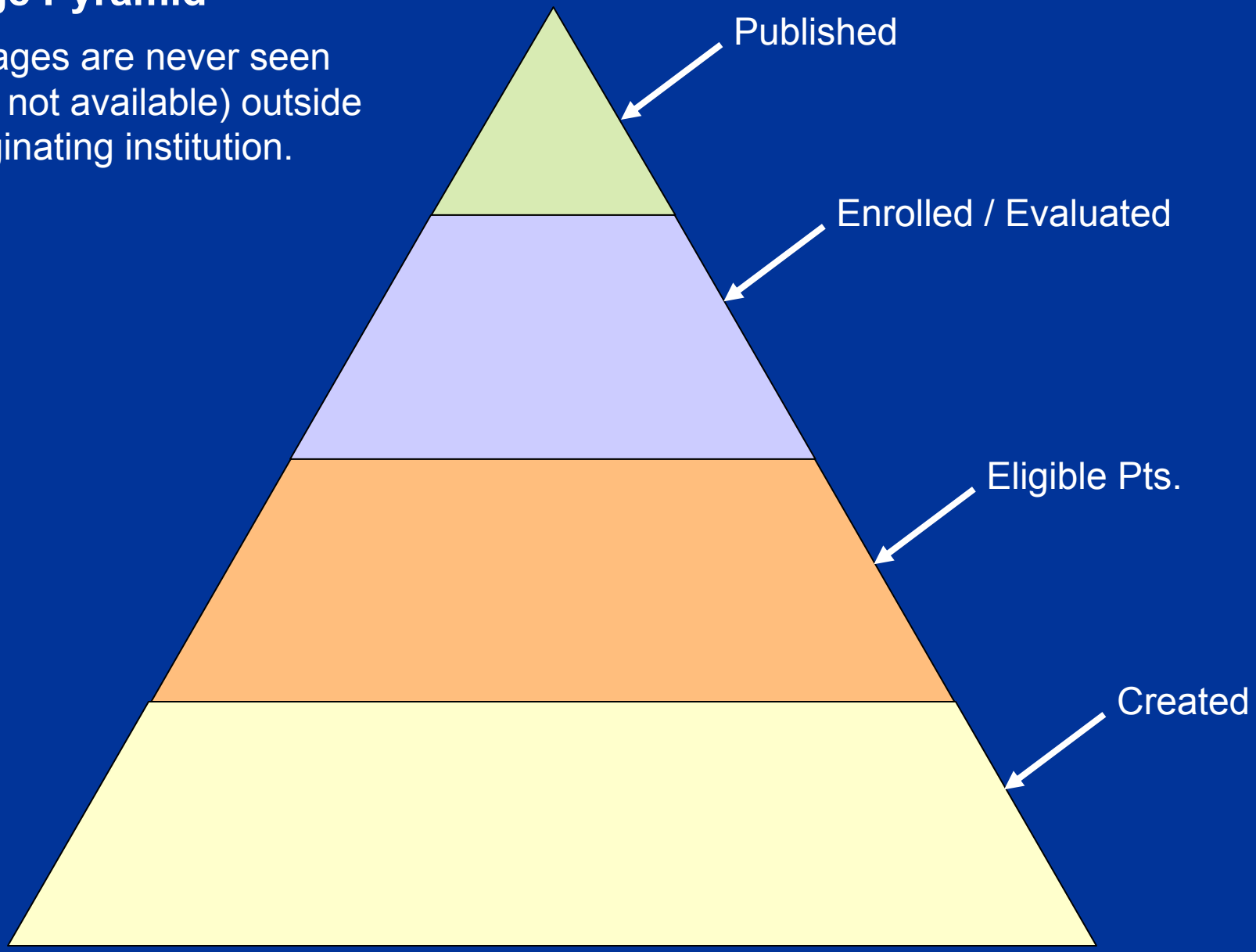


Image Pyramid

Most images are never seen (and are not available) outside their originating institution.



Current Informatics Status

- Genomics / Proteomics
  - Public repositories are common
  - Links to primary data are integrated into publications
  - Indexing and retrieval are "free", open and available to anyone
  - There are open source tools to use biological knowledge resources and apply them in investigations
- Imaging
  - Few (?no) public repositories
  - Few links to primary data in publications
  - Primary data is rarely available, even temporarily
  - Few open source tools that operate on images in conjunction with other forms of biological knowledge/databases

INNOVATION

Big physics, small doses: the use of AMS and PET in human microdosing of development drugs

PET = Positron Emission Tomography
AMS = Accelerator Mass Spectrometry

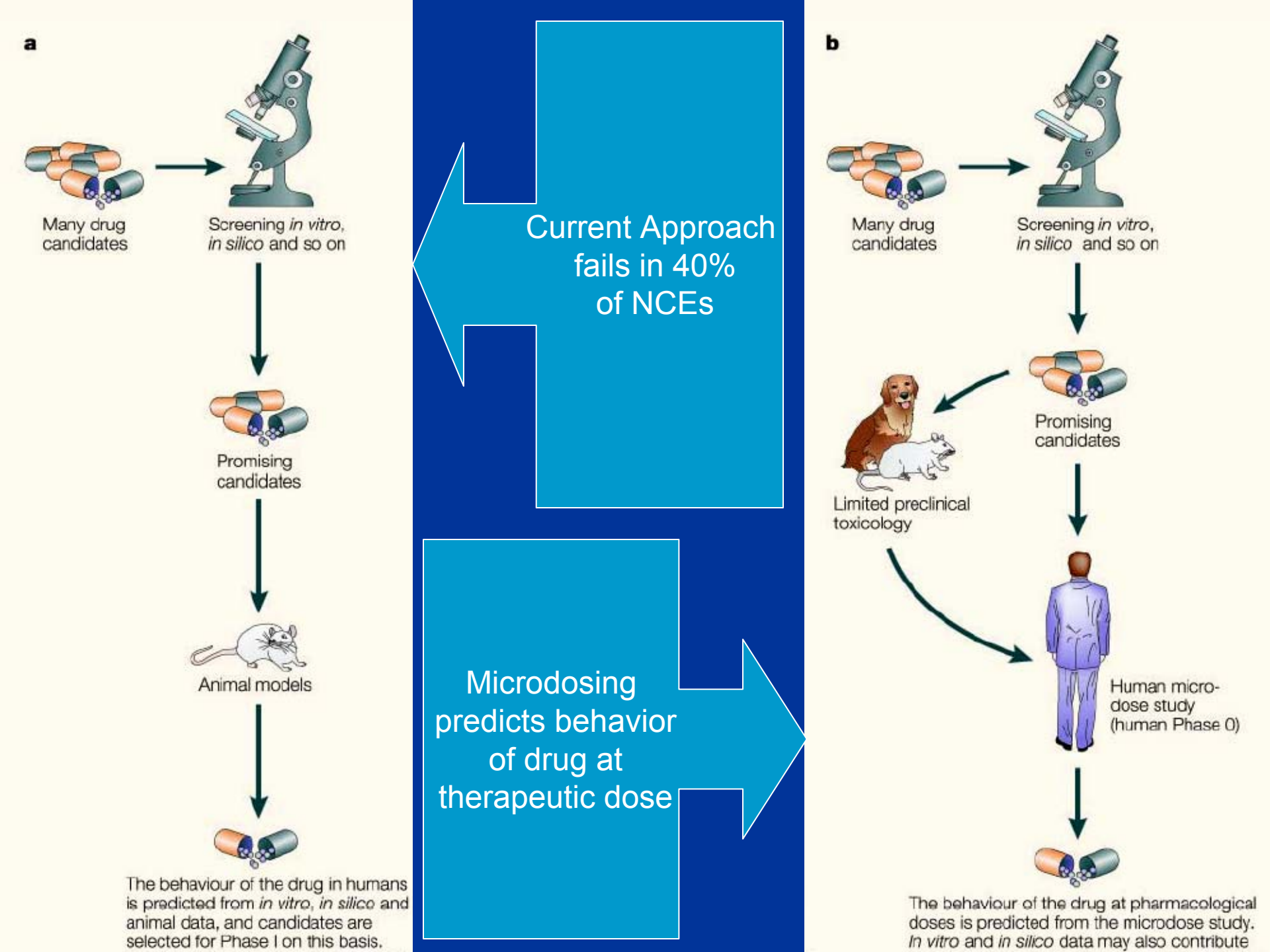
Graham Lappin and R. Colin Garner

The process of early clinical drug development has changed little over the past 20 years despite an up to 40% failure rate associated with inappropriate drug metabolism and pharmacokinetics of candidate molecules. A new method of obtaining human metabolism data known as microdosing has been developed which will permit smarter candidate selection by taking investigational drugs into humans earlier. Microdosing depends on the availability of two ultrasensitive 'big-physics' techniques: positron emission tomography

escalating to in excess of US $800 million per registered drug[1]. Much of this cost is actually associated with those drugs that do not make it to market; therefore the higher the attrition rate, the higher the cost of those drugs that eventually do make it. Unless these costs can be substantially reduced, there will be very few new drugs receiving regulatory approval.

At present, there is an inverse relationship between research and development expenditure by the pharmaceutical industry and the number of drugs receiving regulatory approval. In 2001, only 24 new molecular enti-

NATURE REVIEWS | DRUG DISCOVERY; VOLUME 2 | MARCH 2003 | **233-240**


a
Many drug candidates
Screening *in vitro*, *in silico* and so on
Promising candidates
Animal models
The behaviour of the drug in humans is predicted from *in vitro*, *in silico* and animal data, and candidates are selected for Phase I on this basis.
Current Approach fails in 40% of NCEs
Microdosing predicts behavior of drug at therapeutic dose
b
Many drug candidates
Screening *in vitro*, *in silico* and so on
Promising candidates
Limited preclinical toxicology
Human micro-dose study (human Phase 0)
The behaviour of the drug at pharmacological doses is predicted from the microdose study. *In vitro* and *in silico* data may also contribute

Is proteomics heading in the wrong direction?

Lukas A. Huber

Proteomics is now considered to be one of the most important 'post-genome' approaches to help us understand gene function. In fact, several genomics companies have launched large-scale proteomics projects, and have started to annotate the entire human proteome. The 'holistic view' painted by a human proteome project is seductive, but is it realistic?

"Proteome indicates the proteins expressed by a genome or tissue" — Marc Wilkins, 1994 (BOX 1). Proteomics is therefore any global analysis of changes in the quantities, and post-translational modifications, of all the proteins in cells, taking the genome sequence as a starting point. Growth, differentiation, senescence, environmental changes, genetic collection of proteins that will differ from individual to individual, and even from cell to cell. Although it is meaningful to talk of 'the human genome' as a species-typical set of genes, on the basis of the definition above it is highly unlikely that there will be a single collection of proteins that can be defined as 'the human proteome' — instead, there will be many proteomes that are characteristic of specific cell types and disease states.

Proteomics is the application of evolving technologies (BOX 2) to analyse proteins on a large, 'genomic' scale to study protein-expression profiles — for example, to compare physiological and disease states. These technologies include two-dimensional (2D)-gel electrophoresis, chromatography, mass spectrometry (MS), bioinformatics and protein 'chips'. One of the first challenges for

Problems with proteomics

- The main difference between genomics and proteomics is that the genome is a static collection of genes, whereas the proteome is not a concrete entity, but rather a dynamic collection of proteins that will differ from individual to individual, and even from cell to cell.
- "There will be many proteomes that are characteristic of specific cell types and disease states."

Subcellular localization of proteins

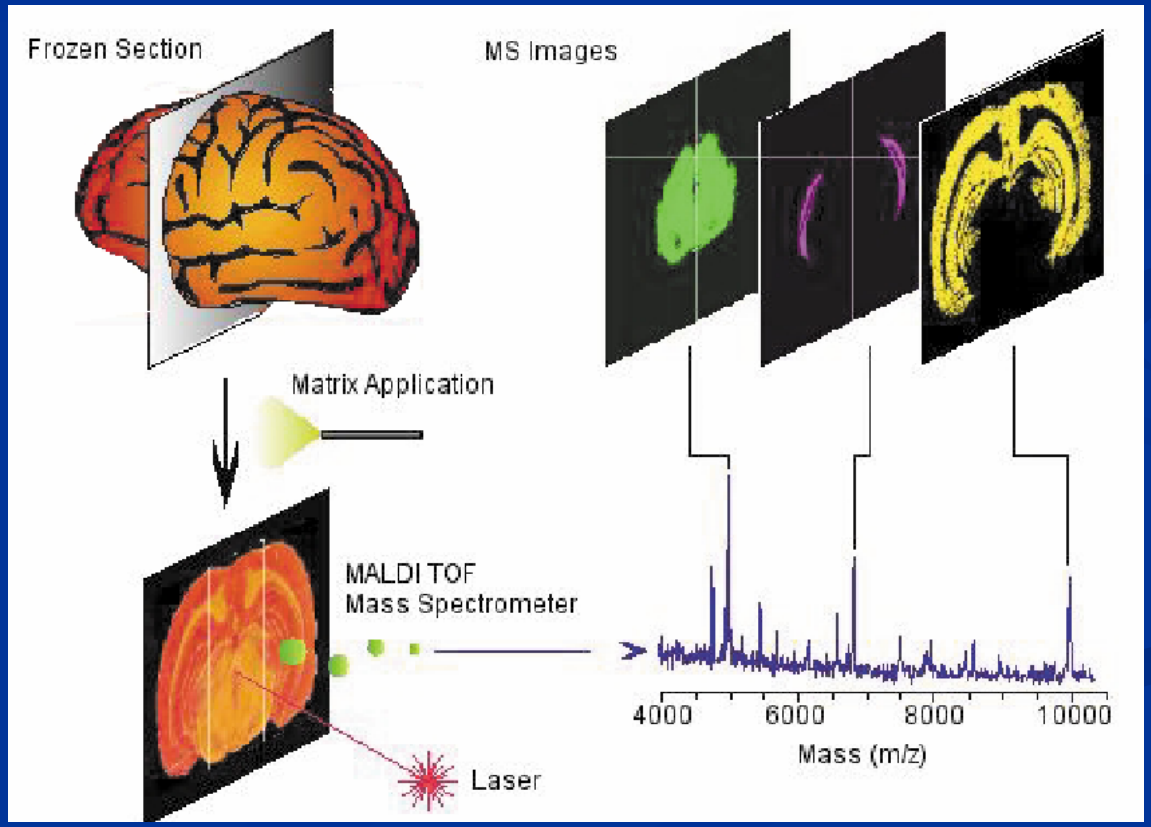
- By confirming the subcellular localization of proteins and their molecular interactions, we can learn a great deal about the functions of proteins — and that, after all, is the whole point of proteomics.

- Subcellular proteomes, protein-interaction networks and large signalling complexes provide unprecedented opportunities to unlock the mysteries of biological processes and to develop new rational therapeutics (proteomics will soon be competing with proven technologies for, for example, target identification and validation in drug discovery).

Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues

MARKUS STOECKLI, PIERRE CHAURAND, DENNIS E. HALLAHAN & RICHARD M. CAPRIOLI
Mass Spectrometry Research Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Nature Medicine – Vol 7, No. 4, Apr 2001; 493-496.



MOLECULAR IMAGING IN DRUG DISCOVERY AND DEVELOPMENT

Markus Rudin and Ralph Weissleder‡*

Imaging sciences have grown exponentially during the past three decades, and many techniques, such as magnetic resonance imaging, nuclear tomographic imaging and X-ray computed tomography, have become indispensable in clinical use. Advances in imaging technologies and imaging probes for humans and for small animals are now extending the applications of imaging further into drug discovery and development, and have the potential to considerably accelerate the process. This review summarizes some of the recent developments in conventional and molecular imaging, and highlights their impact on drug discovery.

Nature Drug Discovery – Feb 2003 – Vol 2; 123-130.

Imaging in Drug Discovery

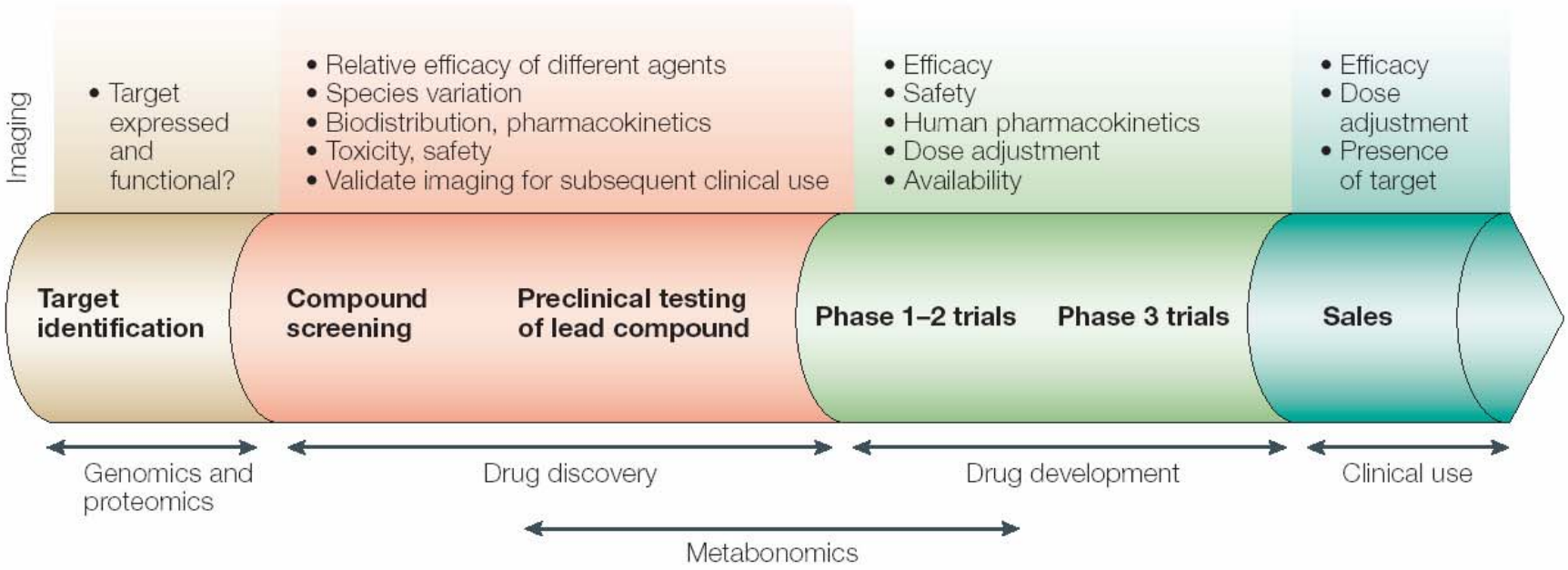
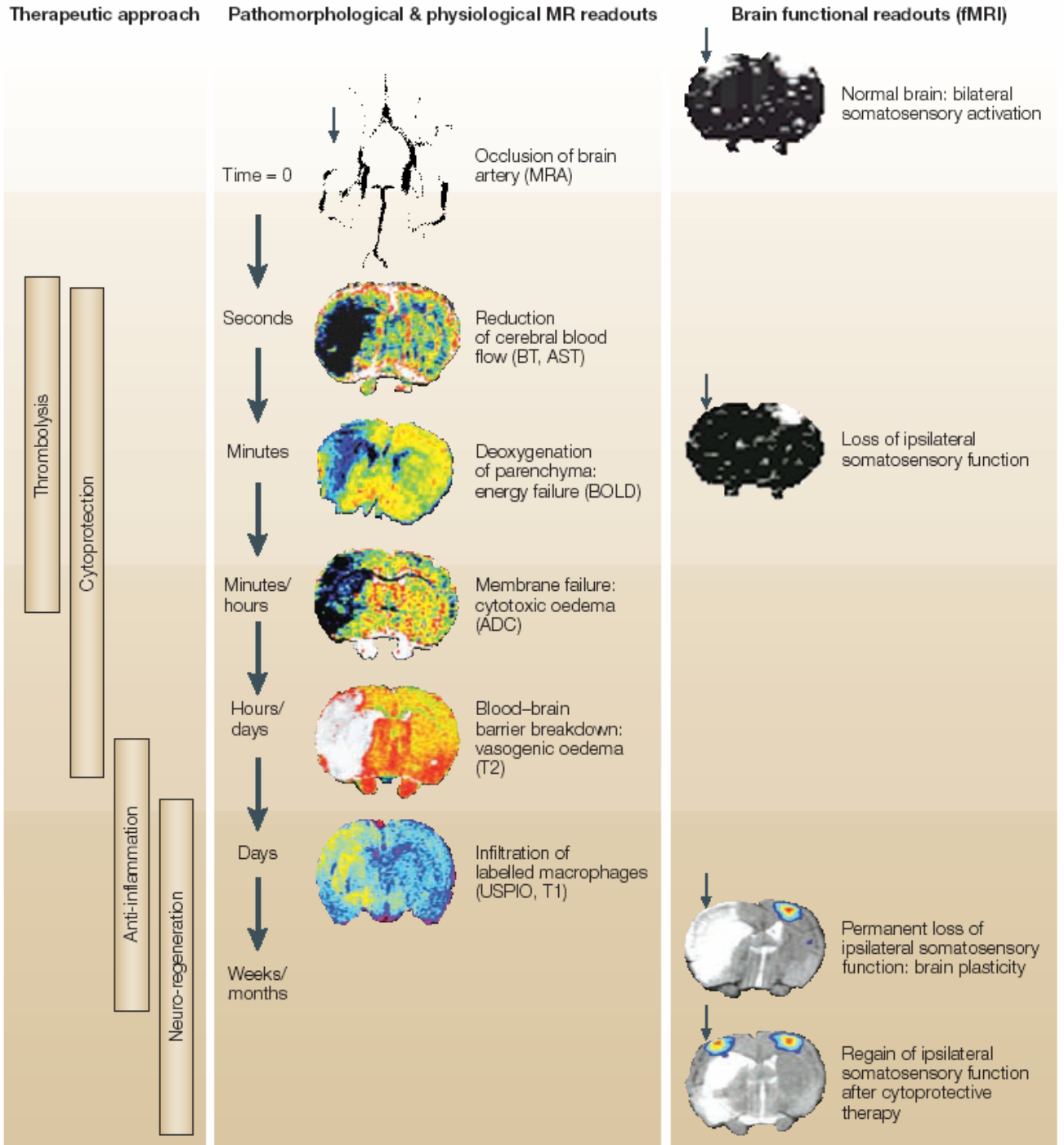


Figure 1 | **Imaging applications in the drug discovery and development process.**

Nature



Structural and Functional Imaging of Small Animals

TIMELINE

New Safe Medicines Faster: a proposition for a pan-European research effort

Ole J. Bjerrum

Providing support for downstream drug development has not traditionally been a primary focus of public research funding programmes. However, the European Commission has decided to include the drug development process in the European Union's Sixth Framework Programme for Research and Technological Development (2002–2006). If the present proposal is ado
mus

trend and address the pressing need for new drugs, a more efficient development process with a larger capacity must be implemented, without in any way compromising safety. Carrying out more research into better methodologies and technologies that could be applied to all stages of the drug development process is the most obvious way to remedy the situation. Properly carried out

Nature Drug Discovery – Vol 1, May 2002, 395-398

the drug development process and thereby bring new, safe and economically affordable medicines to the market. At present, the academic institutions that are carrying out research relevant to the drug development process are small and fragmented, with only weak communication networks. Furthermore, the regulatory agencies do not conduct experimentally based research, which means that they have a conservative approach to the implementation of new methodologies and technologies. The European Agency for the Evaluation of Medicinal Products (EMEA), which is based in London, has not even become fully integrated and operational yet.

The EUFEPS initiative

In view of these circumstances, the Committee of Industrial Relations of the European Federation for Pharmaceutical Sciences (EUFEPS) decided in spring 1999 to approach the European Commission to request funding for research into ways to

Modern biomedical research: an internally self-consistent universe with little contact with medical reality?

David F. Horrobin

Congruence between *in vitro* and animal models of disease and the corresponding human condition is a fundamental assumption of much biomedical research, but it is one that is rarely critically assessed. In the absence of such critical assessment, the assumption of congruence may be invalid for most models. Much more open discussion of this issue is required if biomedical research is to be clinically productive.

bear fruit in time given enough further investment. There is, of course, no way to critically evaluate these two alternatives: all we can do is hope.

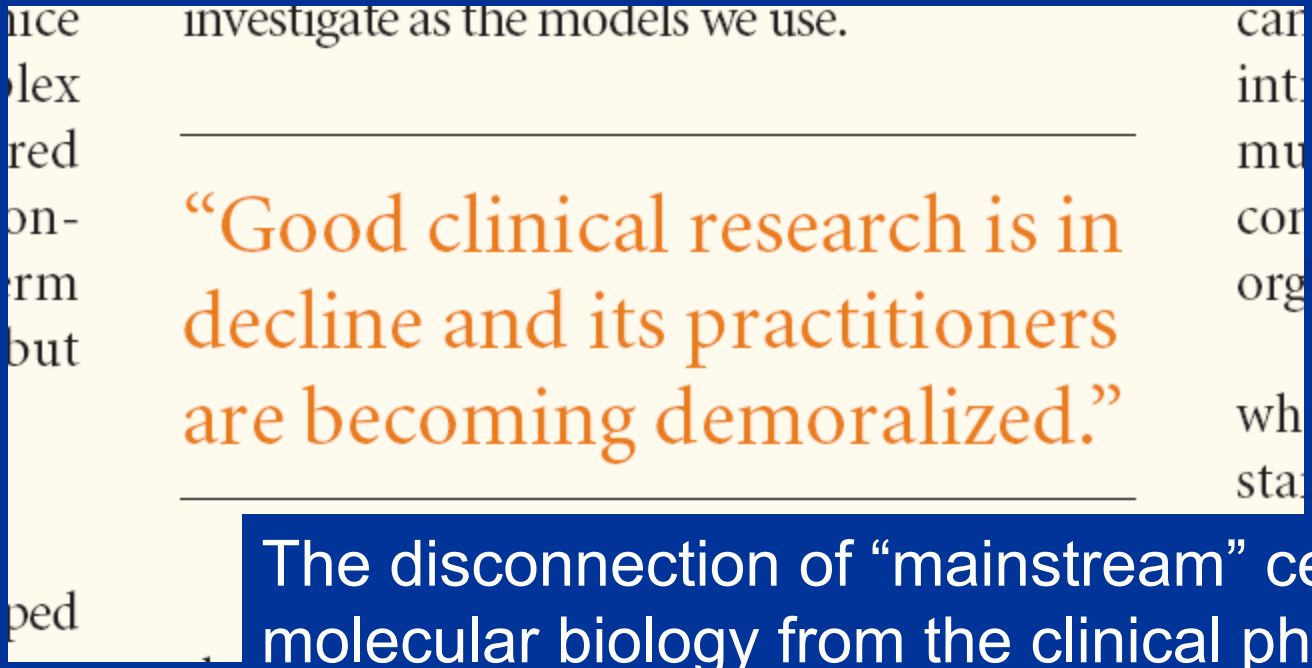
A fourth possibility does deserve critical evaluation, but attracts surprisingly little open discussion. This is that biomedical science, and hence pharmaceutical science, has taken a wrong turn in its relationship to human disease. This discussion paper raises some issues which have been swept under by the pharmaceutical and medi-

Nature Drug Discovery – Vol 2; Feb 2003; 151-154

Box 1 | **Examples of lack of congruence between *in vitro* and *in vivo* models**

- The anatomical constraints and the cellular populations present in culture and *in vivo* are different. There is no circulation *in vitro*.
- The types and rates of nutrient and oxygen supply, and carbon dioxide and metabolite removal, are different.
- The restraints on cell multiplication are different.
- The endocrine environment is different, both in terms of the amounts and patterns of hormones present and their kinetic changes.
- The antibiotic environment is different: *in vivo* cells are not normally bathed in penicillin, streptomycin and other antibiotics, but there has been no systematic evaluation of the effects of any of these exogenous agents on metabolism.
- The lipid environment is different. The phospholipid composition of cells in culture is quite different from the phospholipid composition of the parent *in vivo* cells[2]. As phospholipid composition determines the quaternary structure and therefore function of a high proportion of a cell's proteins, and also determines signal transduction responses to most protein changes, it is likely that the functions of proteins *in vitro* will be, for the most part, somewhat different from the functions of those same proteins *in vivo*[3].
- Even when appropriate constituents are present in culture fluid, their concentrations may be dramatically different from anything seen *in vivo*.

Potential contribution of imaging – Reconciliation of in vivo (human and animal) with in vitro biology


investigate as the models we use.

"Good clinical research is in decline and its practitioners are becoming demoralized."

The disconnection of "mainstream" cell and molecular biology from the clinical phenotype (in vivo human condition) can be resolved to a significant degree by Imaging.

Identification of a gene causing human cytochrome *c* oxidase deficiency by integrative genomics

Vamsi K. Mootha*, Pierre Lepage†, Kathleen Miller*, Jakob Bunkenborg‡, Michael Reich*, Majbrit Hjerrild‡, Terrye Delmonte*, Amelie Villeneuve†, Robert Sladek§, Fenghao Xu¶, Grant A. Mitchell‖, Charles Morin**, Matthias Mann‡, Thomas J. Hudson§, Brian Robinson¶, John D. Rioux*†‡‡‡, and Eric S. Lander*†‡‡‡§§

*Whitehead Institute/Massachusetts Institute of Technology Center for Genome Research, Cambridge, MA 02139; †Genome Quebec Innovation Centre, McGill University, Montreal, QC, Canada H3G 1A4; ‡MDS Proteomics, 5230 Odense, Denmark; §Montreal Genome Centre, McGill University Health Centre, Montreal, QC, Canada H3G 1A4; ¶Hospital for Sick Children, Toronto, ON, Canada M5G 1X8; ‖Service de Génétique Médicale, Hôpital Sainte-Justine, Montreal, QC, Canada H3T 1C5; **Department of Pediatrics and Clinical Research Unit, Chicoutimi, QC, Canada G7H 4A3; and §§Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02138

Contributed by Eric S. Lander, November 25, 2002

Identifying the genes responsible for human diseases requires combining information about gene position with clues about biological function. The recent availability of whole-genome data sets of RNA and protein expression provides powerful new sources of functional insight. Here we illustrate how such data sets can expedite disease-gene discovery, by using them to identify the gene causing Leigh syndrome, French-Canadian type (LSFC, Online Mendelian Inheritance in Man no. 220111), a human cytochrome *c* oxidase deficiency that maps to chromosome 2p16-21. Using four public RNA expression data sets, we assigned to all human genes a "score" reflecting their similarity in RNA-expression profiles to known mitochondrial genes. Using a large survey of organellar proteomics, we similarly classified human genes according to the likelihood of their protein product being associated with the mitochondrion. By intersecting this information with the relevant genomic region, we identified a single clear candidate gene, *LRPPRC*. Resequencing identified two mutations on two independent haplotypes, providing definitive genetic proof that *LRPPRC* indeed causes LSFC. *LRPPRC* encodes an mRNA-binding protein likely involved with mtDNA transcript processing, suggesting an additional mechanism of mitochondrial pathophysiology. Similar

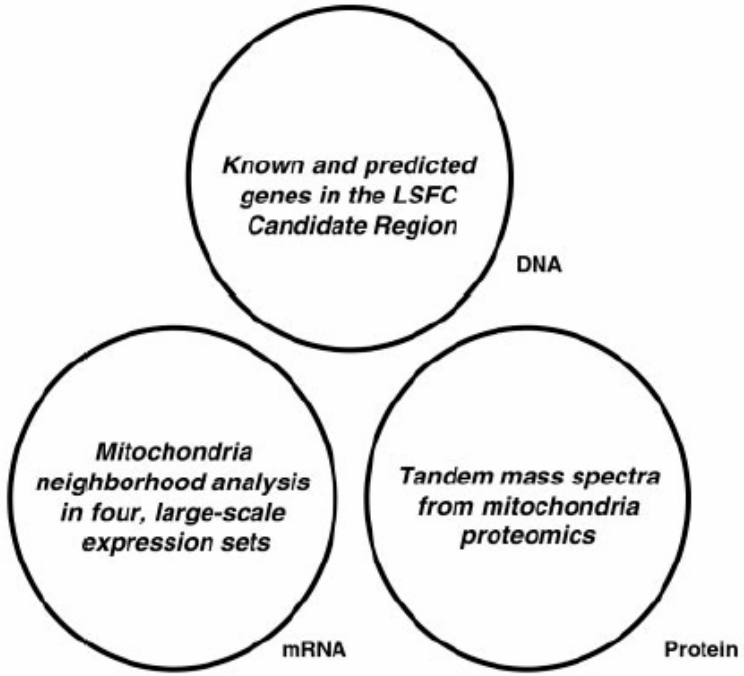


Fig. 1. DNA, mRNA, and protein data sets that are used in this study.

Another opportunity for imaging

- Once a human disease gene has been identified and its sequence variants characterized, this information can be used as a diagnostic tool.
- Nevertheless, details of the biological role(s) of the newly cloned gene are often either sparse or entirely lacking.
- Gene orthologs in other species (e.g., C. elegans) allows understanding the abnormal human gene function in a better understood, simpler animal model.

C. elegans

- Transparent, free-living soil nematode
- 1 mm long, maximum diam ~ 80 μ
- Diet = E. coli
- Short life cycle (3.5 days)
- Large number of progeny (>300)
- 959 somatic cells; 302 are neurons
- Complete cell lineage; complete nervous system connectivity; detailed gene map (> 2000 loci) are known.

© 2000 Oxford University Press
Human Molecular Genetics, 2000, Vol. 9, No. 6 Review 869–877

A role for *Caenorhabditis elegans* in understanding the function and interactions of human disease genes

Emmanuel Culetto and David B. Sattelle+

MRC Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK

Received 28 January 2000; Accepted 7 February 2000

A growing number of medical research teams have begun to explore the experimental advantages of using a genetic animal model, the nematode worm *Caenorhabditis elegans*, with a view to enhancing our understanding of genes underlying human congenital disorders. In this study, we have compared sequences of positionally cloned human disease genes with the *C.elegans* database of predicted genes. Drawing on examples from spinal muscular atrophy, polycystic kidney disease, muscular dystrophy and Alzheimer's disease, we illustrate how data from *C.elegans* can yield new insights into the function and interactions of human disease genes.

Gene orthologs for human diseases.

Based on similarity search (using existing databases)

Examples: Friedreich ataxia, spinal muscular atrophy, polycystic kidney disease; muscular dystrophy;

Nomarski micrographs of several knockdown phenotypes

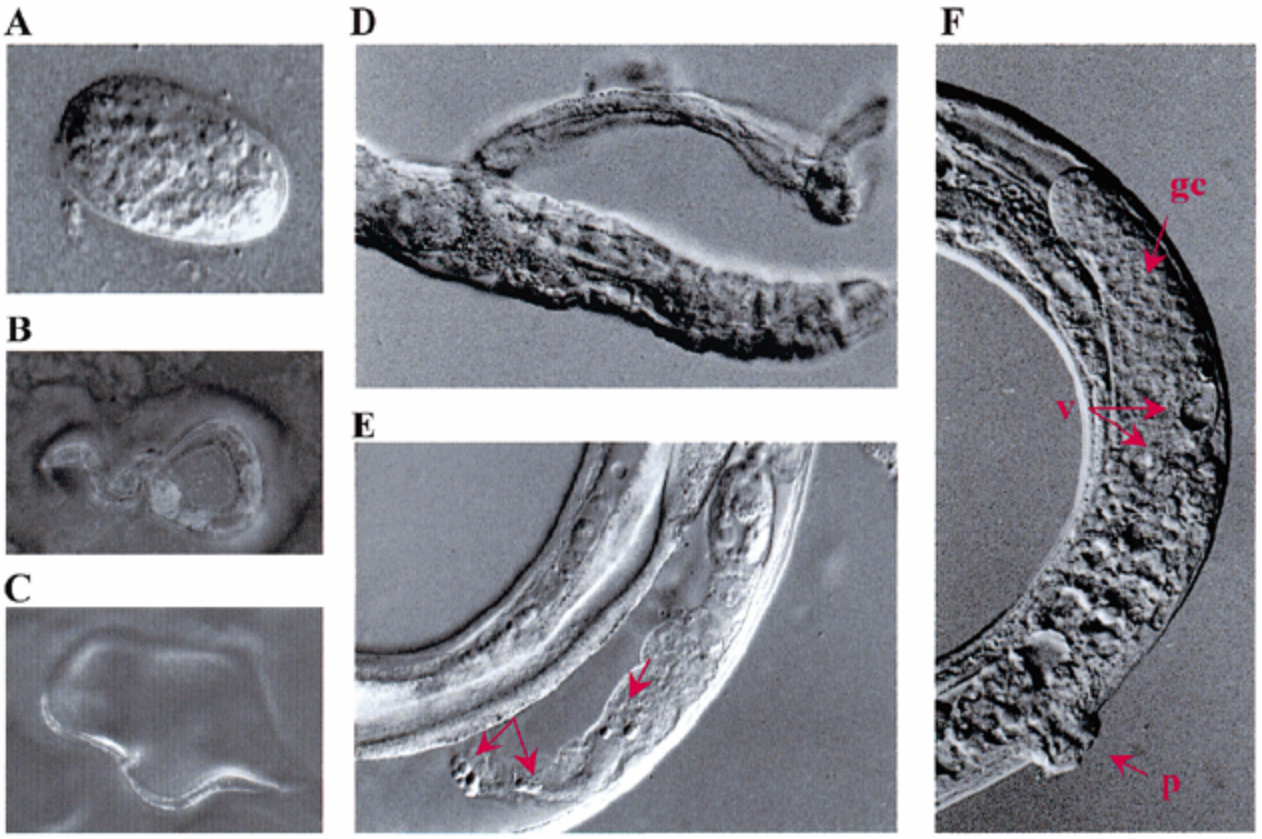
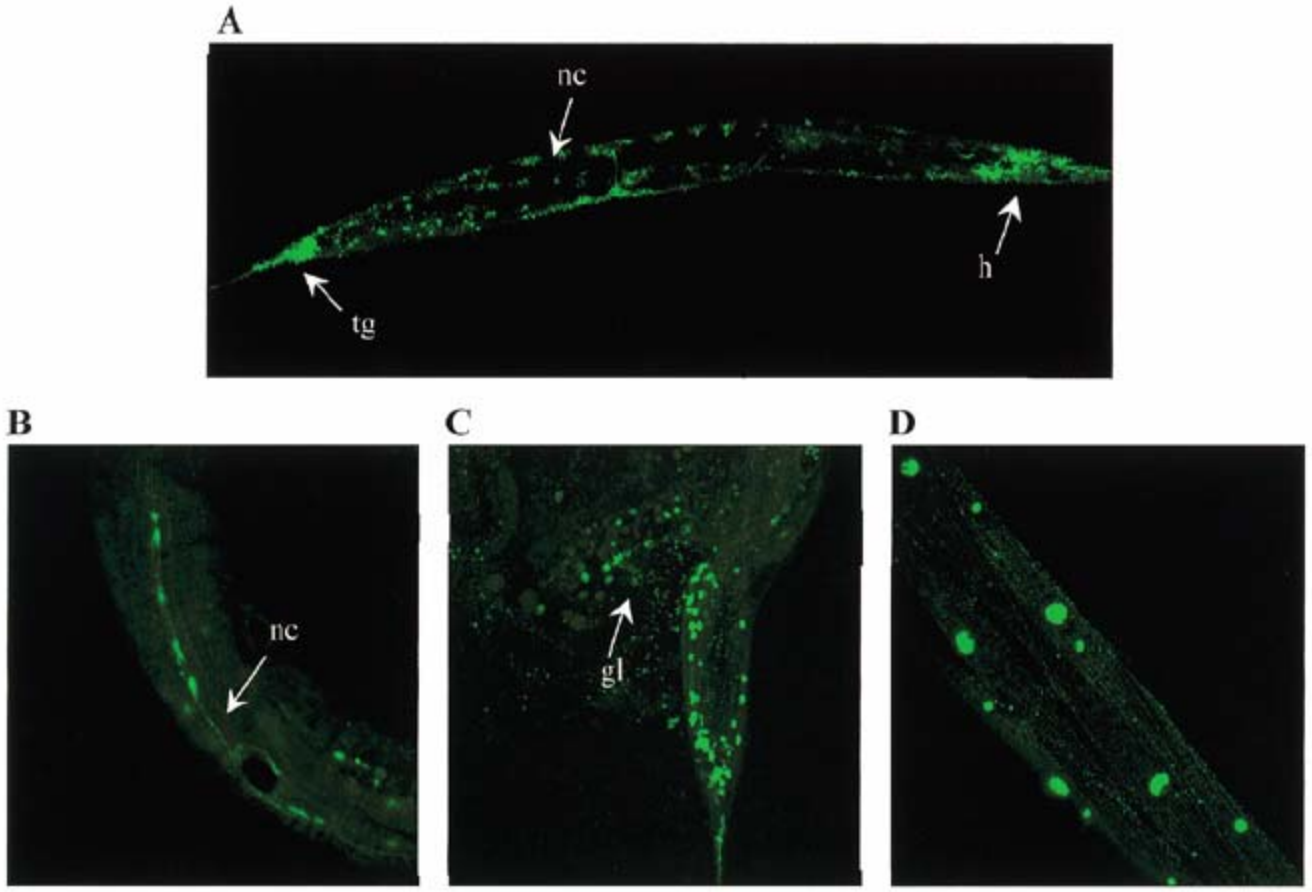


Figure 6. RNA interference of CeSMN. Nomarski micrographs of several knockdown phenotypes. (**A**) Typical arrest phenotype, ~300 min after the fir
The embryonic blastomeres are disorganized and death occurs soon after the completion of gastrulation, presumably due to the inhibition of z

Expression of CeSMN-GFP in transgenic adults



Conclusion

- Imaging is important and will continue to grow.
- Images define phenotypes (normal and abnormal)
- Lack of bioinformatics support for imaging and sequestration of imaging data impairs progress
- Major opportunities - image exploitation, tool and infrastructure development, increased specificity, subcellular location of proteins, open databases – to improve the value of imaging in research

Research needs from imaging

Uncertainty = variability
Quantitative morphology
Subtleties = needle in a haystack
Reconciliation of genotype and phenotype
Optimize systems for tasks
Phase change; transient / dynamic phenomena

nature

27 February 2003 Volume 421 Issue no 6926

Sacrifice for the greater good?

Genome sequencers intend that their community and others should deposit data in community databases immediately, even if it risks the loss of publishing priority. But enforcement of this ideal could be a step too far.

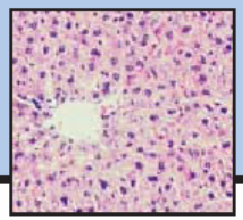
Some US and UK genomics researchers are seeking to extend their principles of open access throughout the world of biology in unprecedented fashion. They claim that enforcing these principles would be in the best interests of science, and they may be right. But anybody believing that researchers in other disciplines and c

dandy for the grandees of the field, who have little to lose in funding and reputation, but are tough on others who are less favoured.

The proposals are long on supporting data release, backed by funding-agency pressure, but short on any means of enforcing the other central requirement: that everyone should behave honourably

data get due credit. *Nature* is all too

This Week



PAGE 995

Interfering with hepatitis

NEWS

SCIENCE PUBLISHING

The UPSIDE of Good Behavior: Make Your Data Freely Available

Prompted by what one biologist calls "an erosion of traditional standards"—and, in particular, the public dispute over access to data on the human genome—leaders in the life sciences have prescribed rules for the sharing of

findings, including source code for software, and they must explain how materials—even those under patent—can be obtained. Authors must be willing to share all this with investigators "on similar, if not identical, terms."

Leader
jected, inc
the U.S. N
Institute i
funded a
Collins wa
dards were
damaging

When
Science h
DNA-sequ
Kennedy,

Science
14 Feb 2003
Vol 299, p. 990.